Faculdade de Engenharia da Universidade do Porto

INVESTIGATIONS ON SURFACE DAMAGE BY ROLLING CONTACT FATIGUE IN ELASTOHYDRODYNAMIC CONTACTS USING ARTIFICIAL DENTS

THE INFLUENCE OF GREASE COMPOSITION

Vitor Manuel Monteiro Borges da Mota

A thesis submitted in conformity with the requirements for the Degree of
Master of Science in Industrial Maintenance,
Department of Mechanical Engineering and Industrial Management,
Faculty of Engineering, University of Porto, Portugal

December, 2005

Abstract

Rolling contact fatigue stands as the main cause limiting the lifetime of machine elements such as bearings, gears or cams. This type of failure is influenced by several parameters among which the "type of lubricant" appears as one of the most important because it controls friction and wear in a given system and also contributes to protect it against contact fatigue. The two basic types of lubricants that are used to lubricate these mechanical elements are oils and greases. In spite of certain limitations in performance, greases are widely used since they provide a low friction, are easily confined and have a long lubricating life at a low cost. However, little knowledge exists on the mechanisms of grease lubrication and particularly on the influence of grease composition on rolling contact fatigue damage. Given this fact, the main motivation for this thesis and its main goal is to understand the influence of grease composition on rolling contact fatigue.

Greases are basically composed of a base-oil, a thickener and in some cases additives. A variation on the values of these components is expected to have consequences on contact fatigue. In order to investigate them, an experimental work was carried out in a twin disc machine using AISI 52100 steel discs (steel commonly used in rolling bearings). Three greases with different compositions and correspondent base-oils were tested. To localize and accelerate the damage by contact fatigue a technique using artificial dents was used.

To evaluate the evolution of the residual stresses due to the case hardened discs, measurements using an X-ray technique were performed in the surface of the discs before and after the experimental tests.

To investigate the influence of a dent on rolling contact fatigue, a numerical simulation using the finite element method was performed and the residual stress and strain fields due to an artificial dent were evaluated.

Finally, stress intensity factors for a surface-breaking crack initiated in an artificial dent were calculated. The applied method used considers the influence of the Hertzian pressure moving along the contact, the lubricant pressure inside the crack and the previously evaluated residual stresses field.

The research presented in this thesis was performed in the scope of the Industrial Maintenance Master Degree and was carried out between September 2004 and December 2005 at the Department of Mechanical Engineering and Industrial Management (DEMEGI) of the Faculty of Engineering of the University of Porto (FEUP).

Resumo

A avaria por fadiga de contacto é uma das principais causas a afectar a vida de elementos mecânicos como os rolamentos, as engrenagens ou as cames. Vários parâmetros a influenciam, sendo que o "tipo de lubrificante" usado aparece como sendo um dos mais importantes, dado que é responsável pelo controlo do atrito e do desgaste num determinado sistema, contribuindo ainda para o proteger da ocorrência de fadiga de contacto. Os óleos e as massas são os dois principais tipos de lubrificantes mais usados na lubrificação de elementos mecânicos. Particularmente as massas, apesar de algumas limitações em termos de desempenho, são amplamente usadas porque garantem um baixo coeficiente de atrito, são facilmente confináveis e possuem uma vida longa a um baixo custo. Contudo, e apesar destas vantagens, existe pouco conhecimento sobre os mecanismos que regem a sua lubrificação e particularmente sobre a influência da sua composição nos fenómenos de fadiga de contacto. É com base neste facto que reside a principal motivação para esta tese. O seu objectivo principal consiste precisamente em compreender qual é a influência da composição das massas lubrificantes na ocorrência de avarias por fadiga de contacto.

As massas lubrificantes são basicamente compostas por um óleo base, um espessante e eventualmente por aditivos. A variação em quantidade destes componentes poderá ter influência na deterioração por fadiga. Para investigar esta possibilidade, foi realizado um trabalho experimental numa máquina de discos usando discos fabricados em aço AISI 52100 (um aço muito comum no fabrico de rolamentos). Três massas lubrificantes com diferentes composições e os respectivos óleos base foram testados. Para localizar e acelerar o dano por fadiga foi usada uma técnica baseada em indentações artificiais.

Para avaliar a evolução das tensões residuais devido ao endurecimento dos discos por tratamento térmico, foram realizadas medições na superfície dos discos, antes e no fim dos ensaios, usando uma técnica de raios X.

Para avaliar a influência das indentações na fadiga de contacto foi realizada uma simulação numérica usando o método dos elementos finitos. Foram determinadas as tensões e as deformações residuais resultantes de uma indentação.

Finalmente, foram calculados os factores de intensidade de tensão para uma fenda iniciada numa indentação artificial. O método usado considera a influência da pressão de Hertz que circula no contacto, da pressão do lubrificante no interior da fenda e da distribuição de tensões residuais previamente determinada.

A pesquisa apresentada nesta tese foi realizada no âmbito do Mestrado em Manutenção Industrial e foi levada a cabo entre Setembro de 2004 e Dezembro de 2005 no Departamento de Engenharia Mecânica e Gestão Industrial (DEMEGI) da Faculdade de Engenharia da Universidade do Porto (FEUP).

Acknowledgements

First, I would like to express my sincere gratitude to my supervisor Assoc. Prof. Luis António de Andrade Ferreira, for his excellent guidance and cooperation during the course of the work.

Among those that made this work to be possible I would like express my gratefulness to:

- Mr. José Fernando Saldanha for all his availability, technical advices and valuable help in redesigning the twin-disc machine so that it was possible to run experiments using greases.

- Eng. Miguel Ferreira from the national oil and gas company GALP Energia for providing the lubricants that were used for the tests.

- the CETRIB collaborators: Eng. Ramiro Martins for his help on the roughness measurements, Eng. Beatriz Graça for the oil viscosity and specific gravity measurements and to Prof. Jorge Seabra and Eng. Luís Magalhães for their advices and help concerning the experimental tests.

- Prof. Castanhola Batista from FCTUC for the residual stresses measurements performed on the discs using the X-ray technique.

- Dr. Carlos Moreira de Sá and Dr. Daniela Silva from CEMUP for the sub-surface analysis using SEM.

- My friend and colleague Eng. Pedro Moreira for his precious help on the numerical simulations using the FEM.

- Mrs. Fernanda Fonseca, secretary of the Applied Mechanics Section office of DEMEGI, for her availability and help in logistics.

Finally I would like to thank to Emese, my mother Isabel, my sister Ana, my father José, all my family and friends for all the, sometimes invisible but always present, encouragement and support.

Acronyms

AISI	American Iron and Steel Institute
ASTM	American Society for Testing and Materials, USA
CEMUP	Centro de Materiais da Universidade do Porto, Portugal
CETRIB	Unidade de Tribologia e Manutenção Industrial, DEMEGI/FEUP
DEMEGI	Departamento de Engenharia Mecânica e Gestão Industrial, FEUP
DF	Departamento de Física, FCTUC
EHD	Elastohydrodynamic
FCTUC	Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Portugal
FEM	Finite Element Method
FEUP	Faculdade de Engenharia da Universidade do Porto, Portugal
INSA	Institut National des Sciences Appliquées, Lyon – France
LEFM	Linear Elastic Fracture Mechanics
LET	Laboratório de Ensaios Tecnológicos, DEMEGI/FEUP
NLGI	National Lubricating Grease Institute
RCF	Rolling Contact Fatigue
SEM	Scanning Electron Microscopy
SIF	Stress Intensity Factor

Contents

List of figures

List of tables

General introduction

Rolling pairs such as bearings, gears or cams are vital elements of machines and devices. They have a direct influence on the reliability and lifetime of mechanical systems. Therefore, it is of major importance recognizing the failure modes of these elements as well as having a complete understanding of the parameters that influence them.

Among the failure modes, *rolling contact fatigue* (RCF) stands as the main cause limiting the lifetime of these elements. The parameters that influence this phenomenon are the operating conditions such as the rolling speed, load, sliding, contact geometry, type of lubricant, temperature, etc. Among these, one of the most important parameters is the type of lubricant because it controls friction and wear in a given system. It also contributes to protect the elements against failure by contact fatigue.

The two basic types of lubricants that are used to lubricate these mechanical elements are oils and greases. Greases are not fundamentally different from oils. In fact, they consist of a base oil (mineral or synthetic) that is trapped in minute pockets formed by soap fibres which constitutes the internal structure of the grease.

In spite of certain limitations in performance, greases are widely used because they provide a low friction, are easily confined and have a long lubricating life at a low cost. It would then be expected that a wide range of research has been carried out an almost, if not complete, understanding on grease lubrication gathered. But, when comparing the current knowledge on the mechanisms of oil and grease lubrication, one concludes that only a little knowledge exists on the mechanisms of grease lubrication. Furthermore, it is still not clear how does grease composition influences the failure modes and particularly RCF.

Given this last fact, the main motivation for this thesis and its main goal is to understand the influence of grease composition on rolling contact fatigue.

The thesis is divided into six chapters. The first two chapters present the context in which the thesis is placed and the remaining four present the experimental and numerical analysis that was carried out.

Chapter I present a general literature review on the RCF damages in elastohydrodynamic contacts. The damages by RCF are characterized, the influent parameters are assigned and the models that were developed by several authors throughout the years are described. In the end of the chapter a technique that uses artificial defects (dents) to study RCF is presented.

Chapter II presents a literature review on oil and grease lubrication with an especial attention to this last one. Given the fact that greases are a main subject

of this work, it is important to give a complete overview including a description on the basic components, lubrication mechanism, characteristics and classification of greases.

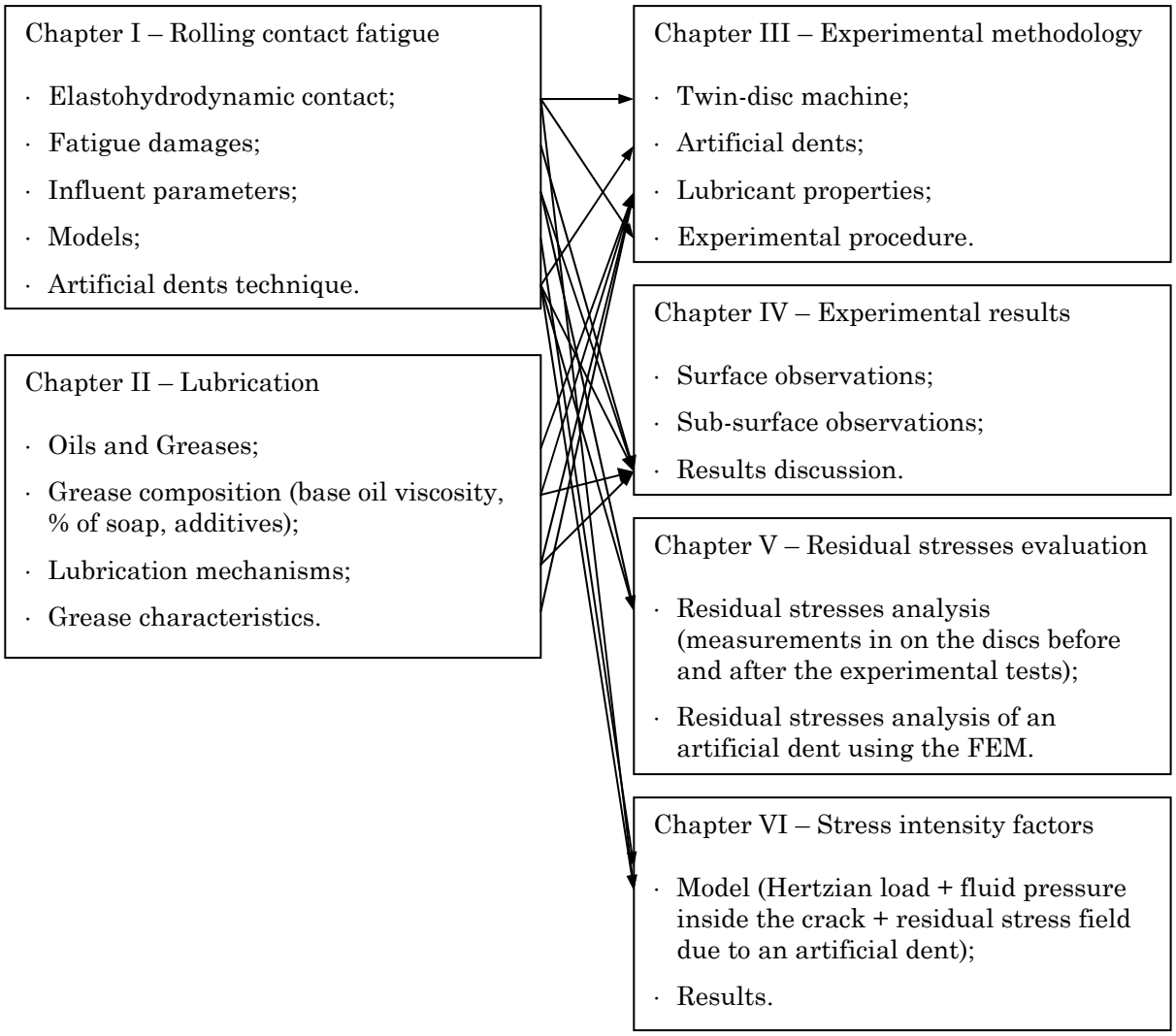
Chapter III presents the experimental methodology that was followed to study RCF damage for the different greases and base oils. Here, a description of the experimental apparatus used (twin-disc machine), characterization of the specimens (discs), a study to choose the most efficient type of artificial dent and the experimental procedure (operating conditions and contact severity assessment) are presented.

Chapter IV presents the experimental results obtained from the tests in the twin-disc machine. A surface and sub-surface analyses are presented. From the results obtained, the hypotheses on the influence of grease composition on rolling contact fatigue are listed.

Chapter V presents the results from the residual stresses measurements performed in a disc in depht, before and after the experimental tests, to study how the residual stresses distribution evolutes. Also a numerical study, using the finite element method (FEM), to evaluate the residual stress field due to an artificial dent (similar to the one used in the experimental tests) is presented.

Chapter VI presents a calculation of the stress intensity factors (SIFs) for a surface-breaking crack initiated in an artificial dent. The applied method considers the influence of the Hertzian pressure moving along the contact, the lubricant pressure inside the crack and residual stresses field due to a dent evaluated in Chapter V. The idea is to study how these different parameters influence the SIFs, and particularly the residual stresses due to a dent, validating the use of this technique as a way of accelerating fatigue damage by rolling contact.

The following scheme illustrates the most important relations that can be established between the most important topics of the literature review (Chapters I and II) and the experimental and numerical work that was carried out in the remaining chapters (Chapters III, IV, V and VI).


Chapter I – Rolling contact fatigue
· Elastohydrodynamic contact;
· Fatigue damages;
· Influent parameters;
· Models;
· Artificial dents technique.
Chapter II – Lubrication
· Oils and Greases;
· Grease composition (base oil viscosity, % of soap, additives);
· Lubrication mechanisms;
· Grease characteristics.
Chapter III – Experimental methodology
· Twin-disc machine;
· Artificial dents;
· Lubricant properties;
· Experimental procedure.
Chapter IV – Experimental results
· Surface observations;
· Sub-surface observations;
· Results discussion.
Chapter V – Residual stresses evaluation
· Residual stresses analysis (measurements in on the discs before and after the experimental tests);
· Residual stresses analysis of an artificial dent using the FEM.
Chapter VI – Stress intensity factors
· Model (Hertzian load + fluid pressure inside the crack + residual stress field due to an artificial dent);
· Results.

Chapter I

Rolling contact fatigue damages in elastohydrodynamic contacts

1. Introduction

In this chapter, a literature review on the study of rolling contact fatigue in elastohydrodynamic contacts is presented.

Machine elements such as gears, rolling bearings or cams (Figure 1.1), may suffer from a number of failure modes, including [1]:

(1) Wear: which is aggravated by severely starved contacts or by lubricant contamination;

(2) Galling: similar to scuffing or scoring of gear teeth, which occurs in contacts where there is a high sliding speed. Insufficient or non-existent lubrication results in material transfer between the surfaces;

(3) Localized plastic deformation: caused by wear debris being drawn into the conjunction;

(4) Overloading: causing bulk plastic deformation of the bodies at the contact;

(5) Overheating: from an outside source or from the contact itself. The effectiveness of the elastohydrodynamic is reduced by a high temperature. Excessive temperatures can cause the surfaces to be softened, eventually leading to gross failure.

(6) Poor assembly of the machine elements: leading, for example, to edge loading.

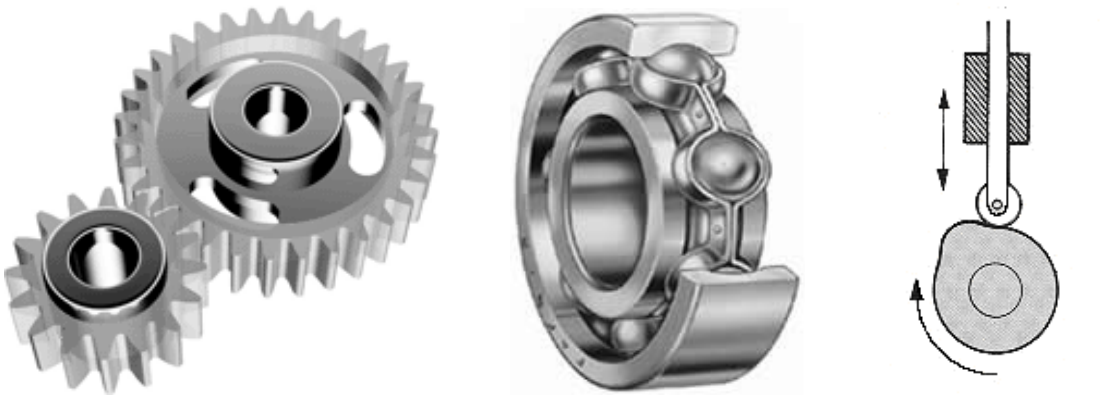


Figure 1.1 – Machine elements with rolling members.

Most of the above mentioned causes can be avoided by careful system design (considering also the lubrication aspects), manufacture and proper maintenance. Nevertheless, as Tallian concludes [2], even if all of these causes are avoided, then the rolling contact mechanism will eventually fail by fatigue. The specific term for this type of fatigue is *"rolling contact fatigue" (RCF)* and it stands as the main cause limiting the life of *elastohydrodynamic (EHD)* contacts. In this type of contacts, the surfaces of the contacting bodies are submitted to a form of hydrodynamic lubrication where the elastic deformation of the containing bodies and the changes of viscosity with pressure play fundamental roles.

RCF is induced by a cyclic loading in the contacting surfaces and leads to two scales of damages [3]:

1. *At the scale of the contact (macro-scale)*: characterized by the formation of macroscopic craters in the contact surface (denominated "spalling");
2. *At the scale of roughnesses (micro-scale)*: characterized by the formation of glazed surface areas (denominated "surface distress").

Both these damages appear after a process that includes a phase of initiation and a phase of propagation of a network of cracks.

The studies on RCF begun in the late nineteenth century, associated with the popular use of ball and roller bearings for general industrial applications. With their increased usage and the increasing sophistication of mechanical design, the necessity to predict their endurance capabilities became a major priority [4].

In 1881, Hertz [5], with his pioneering work, developed the mathematical theory for the general three-dimensional contact problem. He established the equations that provide the analytical solutions for the deformations and stresses to which two non-conformal surfaces are submitted when loaded against each other. His theory is the base of contact mechanics.

In 1902, Stribeck tested and confirmed the predictions of the Hertzian theory for rolling bearings. He established a relation between the applied load necessary to produce in it a permanent deformation in a ball and the diameter of the ball. The geometric configuration, the type and the quality of the steel also intervened in this relation by an empirically determined constant. Stribeck was the first to provide a formula to rolling bearing manufacturers so they could design bearings for each specific application [6].

In 1912, Goodman, who initially identified the phenomenon of rolling contact fatigue, was the first to specify the condition of spalling. He developed a procedure for the early detection of damages in bearings and showed that the load capacity of a rolling element of a bearing diminishes with the speed [6].

In the 1930's, bearings life endurance calculation was one of the main problems of industry. By this time, several studies are carried out to understand the phenomenon of RCF and two directions were taken:

- Way (1935): based on his experimental work, Way [7] was the first one to consider that the fatigue *failures may occur at the surface level.* He concluded that the formation pits is the result of cracks started on the surface which propagate under the combined influence of the stresses induced by the contact and the hydraulic pressure due to the lubricant trapped in the cracks.
- Lunberg and Palmgren (1947-1952): more or less at the same time, Palmgren, based on the work of Stribeck, developed over the following years several equations for calculating the life endurance considering that it is *limited by a critical shear stress value under the surface.* He provided an equation of the specific loading that was later used by Stribeck, as a function of the total number of revolutions, of the loading frequency and of a constant. Considering loading as a function of the number of cycles, Palmgren showed that life endurance is determined by the load and, consequently by fatigue. From 1947 to 1952, Lundberg and Palmgren [8] published their theory on the life rating *"L".*

In the most recent years, the EHD theory provided a common base in approaching the highly loaded lubricated contacts. Nowadays, the pressure field on the surface can be calculated from EHD models. Using the theory of elasticity, it is then possible to determine the internal stresses field due to the surface loadings. Finally, the risk of contact failure can be estimated according to two approaches resulting from the two directions above presented. The first one is represented by the so called "Engineering Models" which are based on the work of Lundberg and Palmgren and confine themselves to variables for which extensive experimental (life test) data and field experience exists. An example is the model of fatigue life proposed by Ioannides and Harris [9] which makes use of the macroscopic mechanical properties of the materials and is widely used in the bearings industry. The second approach is represented by the so called "Research Models" which confine themselves to a limited segment of the failure process (e.g., crack propagation only), to which the research is directed; they are often based on limited experimental background and may not attempt to correlate quantitatively with field experience. They are based on the analysis of the mechanisms of cracking by fracture mechanics and dislocations theory approaches and a distinction is made between the phases of crack initiation and propagation.

The first goal of this chapter is to present the context in which RCF is placed that is the EHD contact. In Section 2 aspects involving the EHD contact such as contact stresses, contact geometry, EHD lubrication mechanism and the influent parameters on these aspects are presented.

In Section 3, RCF is defined through the characteristics and damages developing processes that characterize it. The parameters that influence RCF are also outlined.

Section 4 presents the two approaches referred above for the study of RCF that is the "Engineering Models" (based on the characterization of the materials life) and the "Research Models" (concerns the crack opening modes through the phases of cracking initiation and propagation of rolling fatigue).

Finally, Section 5 presents a review on a technique to evaluate the performance of RCF using artificial dents. This technique induces fatigue cracks near the highly stressed defect site avoiding the randomness of the initiated crack locations and allowing a short time study of RCF, which would not be possible using the conventional and more time consuming fatigue tests.

2. Elastohydrodynamic lubrication

EHD lubrication can be defined as a form of hydrodynamic lubrication where the elastic deformation of the containing bodies and the changes of viscosity with pressure play fundamental roles [10]. This type of lubrication operates generally between non-conforming surfaces that can be found in machine elements like gears, rolling element bearings or cams, although it can occur under certain circumstances in the contacts classified as conformal, such as highly loaded journal and pad bearings which have a significant component of contact and bending deformation.

2.1. Contact stresses

As it was referred above, in many engineering applications, such as rolling bearings, gears, cams, seals, etc., the contacting surfaces are non-conformal (Figure 1.2) hence the resulting contact areas are very small and the pressures are very high. From the point of view of machine design it is essential to know the values of stresses acting in such contacts. These stresses can be determined, from the analytical formulae, based on the theory of elasticity, developed by Hertz.

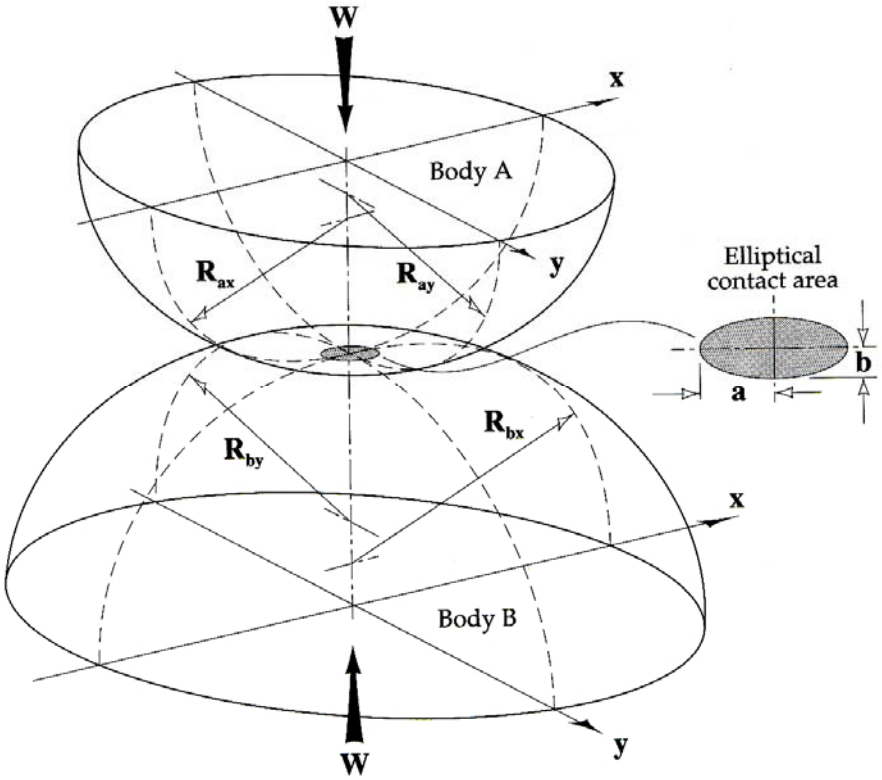


Figure 1.2 – Contact area for a non-conforming contact [10].

Simplifying assumptions to Hertz's theory

Hertz's model of contact stresses forms the basis of the model of EHD lubrication and is based on the following simplifying assumptions [10]:

- the materials in contact are homogeneous and the yield stress is not exceeded,
- contact stress is caused by the load which is normal to the contact tangent plane which effectively means that there are no tangential forces acting between the solids,
- the contact area is very small compared with the dimensions of the contacting solids,
- the contacting solids are at rest and in equilibrium,
- the effect of surface roughness is negligible.

Later, refinements of Hertz's model by different authors have removed most of the assumptions. Even though, Hertz's theory forms the basis of the model of EHD lubrication.

Stress status in lubricated rolling and sliding contacts

The stress field distribution in the contacting bodies depends on the properties of the materials, the roughness of the surfaces, the contact geometry and the ability of the lubricant to generate a lubricant film between the contacting surfaces and it has a direct influence in the occurrence of RCF.

The contact between two bodies in rolling and sliding conditions induces shear stresses which diminish the fatigue life duration of a mechanism. The shear reference stresses usually used are the maximum shear stress τ_{max}, its variation during the cycles or the equivalent Von Mises shear stress, τ_{VM}. In an EHD contact, the maximum shear stresses occur at a certain depth below the surface and depend on the geometry of the contact (spherical or elliptical).

Under pure rolling conditions the shear stress at the surface level will be 0 (see Figure 1.3). But, if sliding occurs between the contacting surfaces, tangential shear stresses at the surface will appear and the maximum shear stress will move from depth to a level close to the surface, as it can be seen in Figure 1.4.

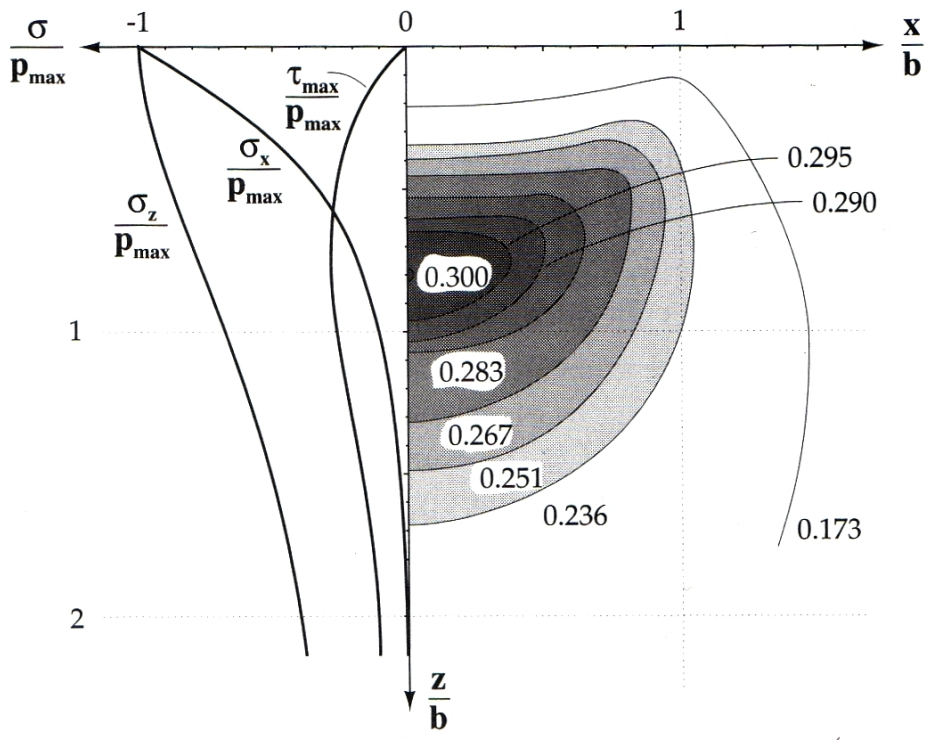


Figure 1.3 – Subsurface stress field for two cylinders in static contact; p_{max} is the maximum contact pressure, b is the half width of the contact rectangle [11].

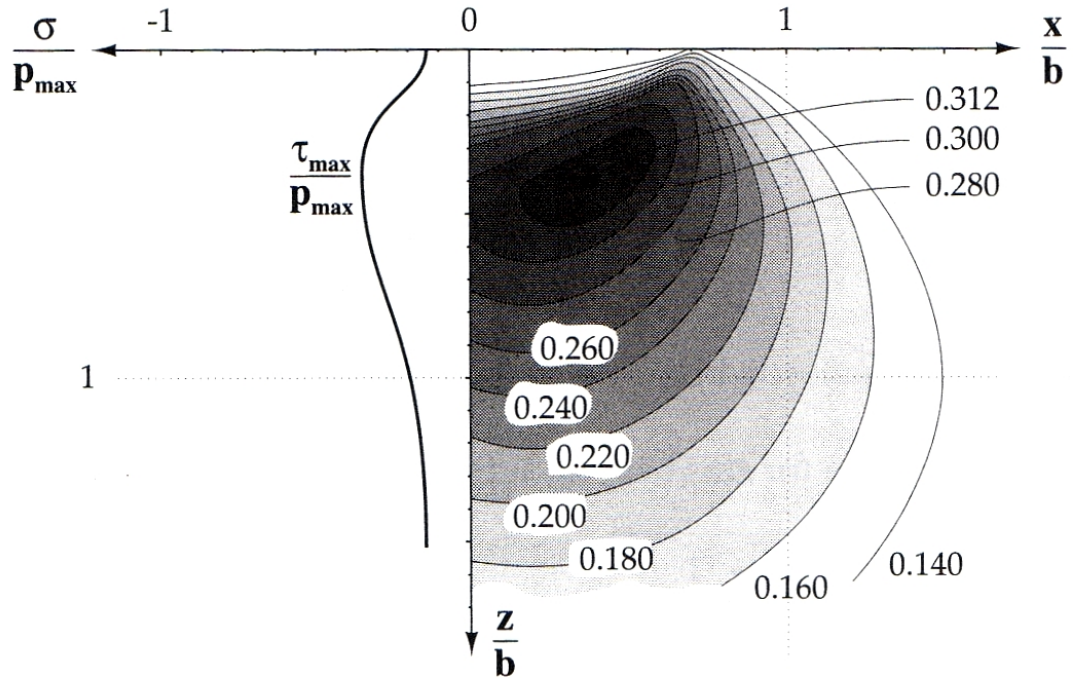


Figure 1.4 – Subsurface stress field for a cylinder sliding on a plane; p_{max} is the maximum contact pressure, b is the half width of the contact rectangle [11].

The influence of the surface roughness in the stress status

In a real contact, the presence of asperities or defects in the surface causes high shear stresses concentration in the surface and in the vicinity. Tallian (see for instance Dumont [6]) showed the influence of the modifications on the lubrication conditions induced by the surface topography, and defined 5 zones in the depth of a body loaded by stresses with different natures and intensities (Figure 1.5).

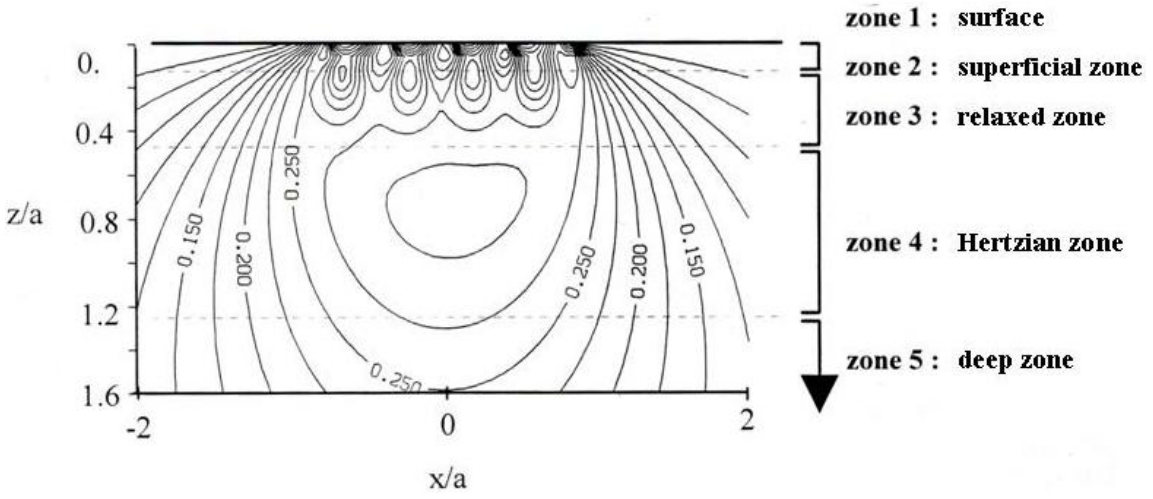


Figure 1.5 – Von Mises shear stress isovalues for a cylinder/plan rough contact [6].

At the surface

At the surface, several defects can exist. The roughnesses effects on the surfaces pressure distribution and on the bodies stress field are also present in the case of the lubricated contacts.

Superficial zone

It is located right under the surface, in a few microns of depth. It is a zone where significant stress concentrations occur, mainly due to the presence of asperities which generate disturbances in the pressure distribution, enhancing significant pressures at the surface level of the solid. This zone is particularly exposed for the initiation of microscopic cracks that may develop to form micro-spalls, accompanied with plastic strains.

Relaxed zone

In this zone, stresses due to pressure disturbances are lower and the Hertzian stresses are not high anymore. The residual stresses generated by coating do not reach this depth (~ 10 μm).

Under pure rolling conditions, from the fatigue analysis point of view, this zone is relaxed. But, the presence of tangential shear stresses in the surface generated by friction can induce a reduction of this zone's depth.

The presence or the absence of this relaxed zone is regarded as particularly significant from the point of view of RCF. It is through it that the propagation of cracks in the Hertzian zone occurs, towards the surface or from the surface towards the subsurface. If this zone is reduced, the propagation will take place more easily.

Hertzian zone

This zone is subjected to the high macroscopic Hertzian stresses. It is centered on the maximum Hertzian alternate shear stress and on the maximum Tresca and Von Mises shear stresses. According to Lundberg and Palmgren, fatigue life is limited by the behavior of this zone. Under normal operating conditions (pure rolling), fatigue starts after a great number of cycles ($>10^8$). The cyclic repetition of the Hertzian stresses induces a mechanism of accumulation of plastic micro-strains and plastic micro-distortions supported by the presence of inclusions in the matrix of the bodies. The rise of micro-dislocations can cause microscopic cracks to appear in the body.

Deep zone

Here, the influence of the Hertzian pressure is not significant anymore. The high stresses come from tight adjustments or centrifugal forces at high rotation speeds.

2.2. Geometry of contacting elastic bodies

Elastic bodies in contact deform and the contact geometry, load and material properties determine the contact area and stresses. The contact geometry depends on whether the contact occurs between surfaces which are both convex or a combination of flat, convex and concave.

The shape of the contact area depends on the shape (curvature) of the contacting bodies. For example, point contacts occur between two balls, line contacts occur between two parallel cylinders and elliptical contacts, which are the most frequently found in many practical engineering applications, occur when two cylinders are crossed, or a moving ball is in contact with the inner ring of a bearing, or two gear teeth are in contact.

To illustrate the contact between two elastic bodies, the simpler contact of tow

parallel cylinders (Figure 1.6) is presented based on the Hertz theory. This type of contact is used to derive contact pressure distribution, stresses and displacements for the two dimensional line contacts. Other contact geometries and resulting stresses can be found for instance in Johnson [6].

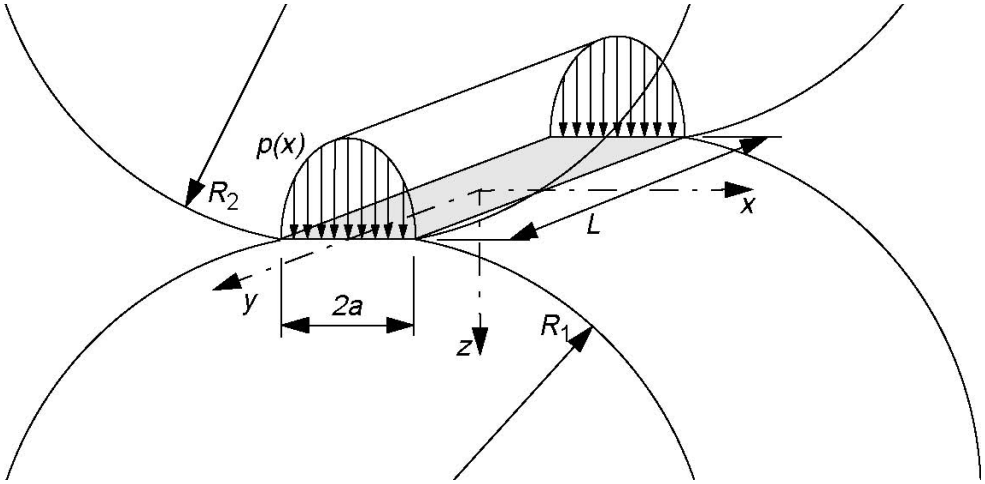


Figure 1.6 – The Hertz pressure distribution of a two dimensional contact [12].

The pressure distribution in the contact region shown in Figure 1.6 is:

$$p(x) = \frac{2P}{\pi a^2 L}\sqrt{a^2 - x^2} \tag{1.1}$$

where P is the force acting over the contact length L and $2a$ is the contact width. The maximum contact pressure at $x = 0$ is

$$P_0 = \frac{2P}{\pi a L} \tag{1.2}$$

The contact width $2a$, the reduced radius of curvature R' and reduced Young's modulus E' are given by

$$\begin{aligned} a &= \sqrt{\frac{4PR'}{\pi L E'}} \\ \frac{1}{R'} &= \frac{1}{R_1} + \frac{1}{R_2} \\ \frac{1}{E'} &= \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \end{aligned} \tag{1.3}$$

The indices in equation (1.3) refer to body 1 and 2, respectively. The normal

stresses σ_x, σ_z and the maximum shear stress σ_x, σ_z and the maximum shear stress, τ_1, on the z-axis ($x = 0$) are drawn in Figure 1.3. The shear stress reaches a maximum, $(\tau_1)_{max} = 0.30p_0$, for $z = 0.78a$. In the literature $(\tau_1)_{max}$ is often referred to as the maximum Hertzian shear stress. Note that the normal stresses along the symmetry line in Figure 1.3 are compressive everywhere. For the line contact, this is the case for every location inside the bodies, except at the surfaces outside the contact, where all stresses are zero.

For an axi-symmetric Hertz contact the contact radius and contact pressure are given by

$$a = \left(\frac{3PR'}{4E'}\right)^{\frac{1}{3}} \text{ and } p(r) = p_0\sqrt{1-\left(\frac{r}{a}\right)^2} \text{ , where } p_0 = \frac{3P}{2\pi a^2} \tag{1.4}$$

The maximum shear stress is $(\tau_1)_{max} = 0.31p_0$, for $z = 0.48a$, when $\nu = 0.3$. In this case the radial surface stress is tensile outside the contact, with the maximum

$$\sigma_{rr} = \frac{(1-2\nu)p_0}{3} \tag{1.5}$$

2.3. Elastohydrodynamic lubricating films

The term elastohydrodynamic lubricating film refers to the lubricating oil which separates the opposing surfaces of a concentrated contact. The properties of the this minute amount of oil, typically 1 μm thick and 400 μm across for a point contact, and which is subjected to extremes of pressure and shear, determine the efficiency of the lubrication mechanism under rolling contact [10].

The three following effects that play a major role in the formation of lubrication films in EHD lubrication are:

- the hydrodynamic film formation,
- the modification of the film geometry by elastic deformation,
- the transformation of the lubricant's viscosity and rheology under pressure.

All these three effects act simultaneously and cause the generation of EHD films.

2.3.1. Pressure distribution in EHD films

The elastic deformation of the contacting bodies results in a central region between the entry and exit with almost parallel surfaces. The two-dimensional Hertzian contact pressure profile is also modified as it illustrated in Figure 1.7. The modifications of the Hertz solution may shortly be summarized as follows [12]:

- the length of the contact increases slightly,
- at the entry region the hydrodynamic pressure is lower than the Hertzian value,
- the opposing surfaces within the contact are almost parallel. The film thickness in this region is described by the central film thickness, h_c,
- the lubricant experiences a substantial rise in viscosity as it enters the contact followed by an equally large but sharp decline at the contact exit. To maintain continuity of the flowing lubricant and to compensate for the loss of viscosity at the contact exit, a constriction is formed close to the exit. The minimum film thickness at the constriction is h_0,
- a pressure peak usually larger than the maximum Hertzian pressure, p_0, is generated just upstream of the constriction. The size and steepness of this pressure peak is dependent on the pressure-viscosity characteristics of the lubricant.
- downstream of the constriction the pressure is lower than predicted by Hertzian theory.

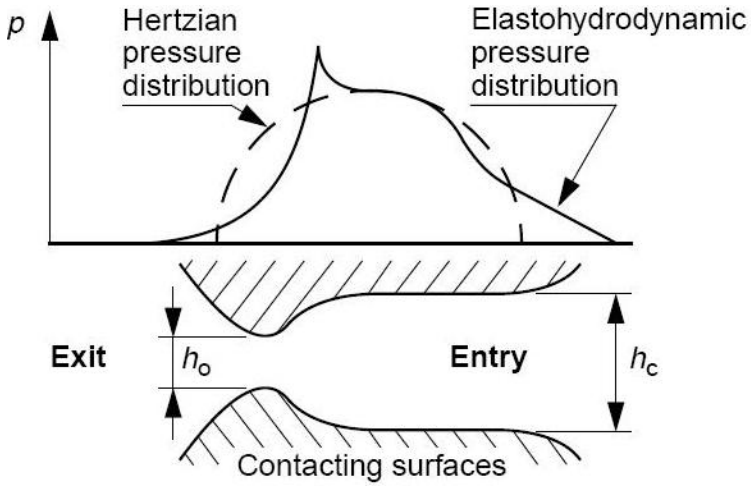


Figure 1.7 – The two dimensional contact pressure under EHD lubrication. The contact moves from the left to the right [10].

The EHD pressure profile is valid for pressures up to 3-4 GPa, which is sufficient for most machine elements [10].

2.3.2. EHD film thickness formulae

The most important information about EHD lubrication was provided by Hamrock and Dowson [13]. The results of their analysis are the formulae for the calculation of the minimum film thickness in EHD contacts. The formulae derived by Hamrock and Dowson apply to any contact, such as point, linear or elliptical, and are now routinely used in EHD lubrication film thickness calculations. They can be used with confidence for many material combinations including steel on steel even up to maximum pressures of 3-4 GPa [13]. The numerically derived formulae for the central and minimum film thickness, as shown in Figure 1.7, are in the following form [13]:

$$\frac{h_c}{R'} = 2.69\left(\frac{U\eta_0}{E'R'}\right)^{0.67}(\alpha E')^{0.53}\left(\frac{W}{E'R'^2}\right)^{-0.067}\left(1-0.61e^{-0.73k}\right) \tag{1.6}$$

$$\frac{h_0}{R'} = 3.63\left(\frac{U\eta_0}{E'R'}\right)^{0.68}(\alpha E')^{0.49}\left(\frac{W}{E'R'^2}\right)^{-0.073}\left(1-e^{-0.68k}\right) \tag{1.7}$$

where:

- h_c is the central film thickness [m];
- h_0 is the minimum film thickness [m];
- U is the entraining surface velocity [m/s], i.e. $U = (U_A + U_B)/2$, where the subscripts *'A'* and *'B'* refer to the velocities of bodies *'A'* and *'B'* respectively;
- η_0 is the viscosity at atmospheric pressure of the lubricant [Pas];
- E' is the reduced Young's modulus [Pa];
- R' is the reduced radius of curvature [m];
- α is the pressure-viscosity coefficient [m^2 / N];
- W is the contact load [N];
- k is the ellipticity parameter defined as: $k = a/b$, where *'a'* is the semiaxis of the contact ellipse in the transverse direction [m] and *'b'* is the semiaxis in the direction of motion [m].

The approximate value of the ellipticity parameter can be calculated with sufficient accuracy from:

$$k = 1.0339\left(\frac{R_y}{R_x}\right)^{0.636} \tag{1.8}$$

where R_x, R_y are the reduced radii of curvature in the *'x'* and *'y'* directions respectively.

The non-dimensional groups in equations (1.6) and (1.7) are frequently referred to in the literature as:

- the non-dimensional film parameter $H = \frac{h}{R'}$
- the non-dimensional speed parameter $U' = \left(\frac{U\eta_0}{E'R'}\right)$
- the non-dimensional materials parameter $G = (\alpha E')$
- the non-dimensional load parameter $W' = \left(\frac{W}{E'R'^2}\right)$
- the non-dimensional ellipticity parameter $k = \frac{a}{b}$

The changes in the non-dimensional parameters have varying effects on the EHD lubrication film thicknesses and pressures. To demonstrate these effects, Hamrock and Dowson allowed one specific parameter to vary while holding all the other parameters constant [13].

2.4. The influence of the surface roughness on EHD lubrication

In the evaluation of EHD film thickness it has been assumed that the contacting surfaces lubricated by EHD films are flat. In practice however, the surfaces are never flat, they are rough, covered by features of various shapes, sizes and distribution. The question arises of how the surface roughness affects the mechanism of elastohydrodynamic film generation [10].

If the surface asperities are of the same height as the EHD film thickness then one may wonder whether there is any separation at all between the surfaces by a

lubricating film. For example, EHD film thickness is often found to be in a range of 0.2 – 0.4 μm which is similar to the surface roughness of ground surfaces [10].

Local film variation as a function of local surfaces roughness is perhaps best characterized by a parameter proposed by Tallian [14]. The ratio of the minimum film thickness to the composite surface roughness of two surfaces in contact is defined as:

$$\lambda = \frac{\Phi_T h_0}{\left(\sigma_A^2 + \sigma_B^2\right)^{0.5}} \tag{1.9}$$

where:

h_0 is the minimum film thickness [m];

σ_A is the RMS surface roughness of body 'A' [m];

σ_B is the RMS surface roughness of body 'B' [m];

λ is the parameter characterizing the ratio of the minimum film thickness to the composite surface roughness;

Φ_T is the thermal effects correction factor given by:

$$\Phi_T = \frac{1 - 13.2\left(\frac{P_0}{E'}\right)L^{0.42}}{1 + 0.213\left(1 + 2.23S^{0.83}\right)L^{0.64}} \tag{1.10}$$

where:

P_0 is the maximum Hertzian contact pressure [GPa];

E' reduced Young's modulus;

S slide-to-roll ratio given by:

$$S = 2\frac{U_1 - U_2}{U_1 + U_2} \tag{1.11}$$

where:

U_1, U_2 linear velocity of the bodies 1 and 2 in the rolling direction;

L the thermal load parameter given by:

$$L = \frac{\eta\beta(U_1 + U_2)}{c} \tag{1.12}$$

where:

- η lubricant's viscosity at the temperature of the surface and local pressure;
- β termoviscosity coefficient;
- c lubricant's specific heat

Figure 1.8 shows the influence of surface roughness on the lubricant film thickness and pressure distribution.

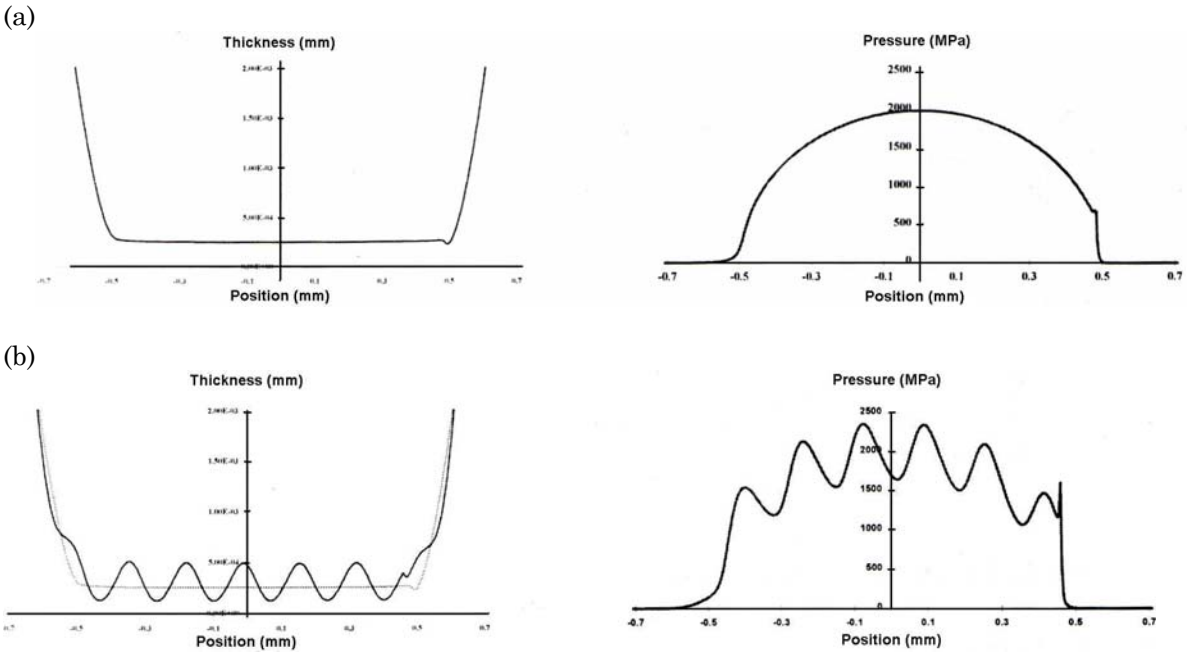


Figure 1.8 – Influence of roughness on the pressure distribution and lubricant film thickness – comparison between a smooth contact (a) and rough contact (b) (adapted from [6])

Measured values of 'λ' have been found to correlate closely with the limits of EHD and the onset of damage to the contacting surfaces such as contact fatigue. It is also possible for wear, i.e. surface material uniformly removed from the contacting surface, to occur when EHD is inadequate. The rapidity of spalling or simple wear is described in the terms of a fatigue life which is the number of rolling/sliding contacts till spalling is sufficient to prevent motion between the opposing surfaces. Figure 1.9 and Table 1.1 present the relationship between 'λ' and surface damages for rolling bearings.

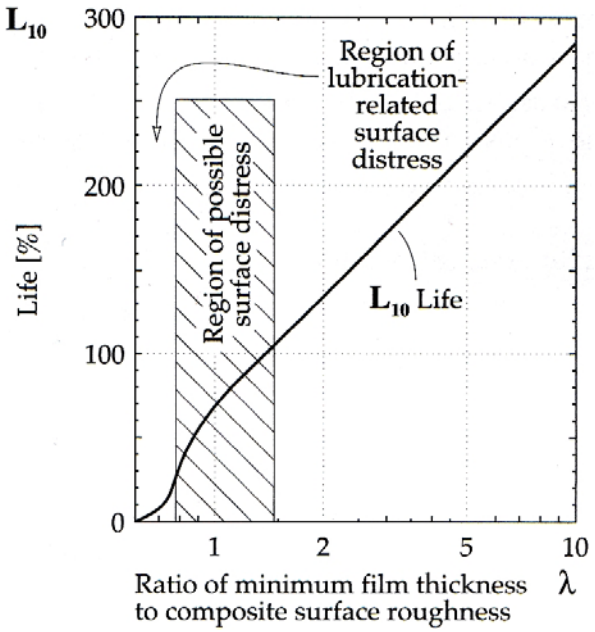


Figure 1.9 – Effects of minimum film thickness and composite surface roughness on contact fatigue life [14]

Table 1.1 – Relationship between 'λ' and surfaces damages for rolling bearings [15].

$\lambda \leq 1.0$	Surface smearing or deformation accompanied by wear can occur.
$1.0 \leq \lambda \leq 1.5$	Surface distress is possible. The term 'surface distress' means that surface glazing and spalling will occur. When the surface has been 'glazed', it is assumed that the original surface roughness has been suppressed by extreme plastic deformation of the asperities.
$1.5 \leq \lambda \leq 3.0$	Some glazing of surface may occur, however, glazing will not impair bearing operation or result in micro-spalling.
$3.0 \leq \lambda \leq 4.0$	Wear can be expected but not glazing.
$\lambda \geq 4.0$	Full separation of the surfaces by an EHL film can be expected.

In EHD 3 regimes of lubrication can be found: 1) full film, 2) mixed film and 3) boundary film. In Table 2 the characteristics of these regimes are presented. In this table the values of λ_0 and λ_1 depend on the type of application (rolling bearings, gears or cams).

Table 1.2 – Regimes of EHD lubrication [15].

λ	Regime	Characteristics
$\lambda \geq 10 \times \lambda_1$	Hydrodynamic	Thick lubricant film separating the contacting surfaces.
$\lambda \geq \lambda_1$	Full film	The contacting surfaces are completely separated by the lubricant.
$\lambda_0 < \lambda < \lambda_1$	Mixed film	The contacting surfaces are partially separated by the lubricant film which allows some contact between the asperities occurs.
$\lambda \leq \lambda_0$	Boundary film	Inexistent lubricant film separating the contacting surfaces and consequently predominant contact between the asperities.

3. Fatigue damage of EHD contacts

The surface damage created by RCF has been given various names in the literature. Following Tallian [3], two scales of damages and three types of damage processes by RCF can be found:

- Scale of damages:
  - Spalling: defined as a macro-scale Hertzian contact fatigue. It is a failure caused by the formation of macroscopic craters in the contact surface as a result of fatigue crack propagation in the Hertzian stress field. When fully developed, the craters are of depth comparable to that of the maximum shear stress (Figure 1.10);
  - Surface distress: defined as micro-scale spalling fatigue. It is a failure of rolling contact metal surfaces by the formation of 1) glazed (burnished) surface areas, 2) asperity scale *micro-cracks* and 3) asperity scale *micro-spall* craters, in that sequence. It is the result of predominantly normal contact stresses acting in asperity dimensions (Figure 1.11).

- Types of damages:
  - Surface origin spalling;
  - Subsurface origin spalling;
  - Surface distress, micro-spalling or micro-pitting.

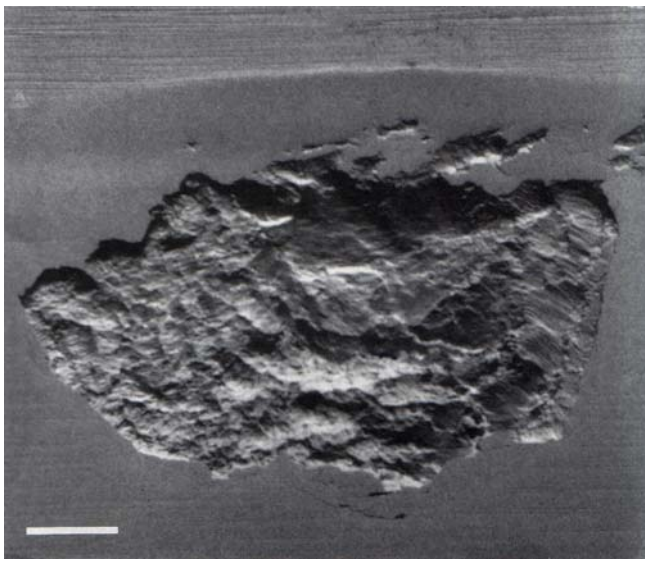


Figure 1.10 – Spall across contact width [3]. Scalebar represents 400 μm.

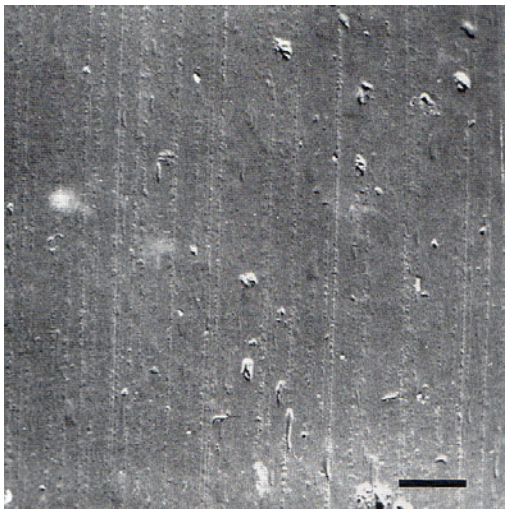


Figure 1.11 – Distressed surface [3]. Scalebar represents 35 μm.

Pitting is a term widely used instead of spalling. It is not truly the appropriate one because it designates the formation of craters instead of fatigue cracking. It is the result of surface porosities due to the manufacture process, corrosion, erosion or a certain type of wear [3].

Concerning surface distress damages a wide vocabulary is used: peeling, fatigue wear, delamination wear, micro-pitting or micro-spalling.

3.1. The general failure process

Although the damage process is not fully understood some events are confirmed [16]. It can be divided into three phases [17]:

1. a brief initial phase of bulk changes in the material;
2. a long, stable phase when only micro-scale changes take place;
3. the final phase, when a macro-crack grows to a spall.

From a crack growth point of view, the third phase may be divided into early crack growth and final fracture. It is, however, assumed that the bulk of the application life is consumed during the first two phases.

During the first phase, bulk changes in the material structure occur in the highly stressed volume under the contact path. Changes may occur in hardness, residual stresses, austenite and martensite structure. The surface roughness is decreased through reduction of both asperity heights and asperity sharpness through wear and local plastic deformation. The surface appearance changes towards that of a

polished surface with smooth low-frequency waviness and scratches remaining from the bottom of the deeper finishing marks.

Alternating micro-plastic flow occurs in the highly stressed volume under the contact path deformation bands during the stable phase which occupies most of the fatigue life of the part. It generates several types of structural and micro-stress alterations. The changes in the material structure brought about by the micro-plastic flow have been described as white etching areas, when disseminated in the highly stressed volume. When created around a defect such as an inclusion or an asperity they may be designated as butterflies. Sometimes during this stable phase micro-cracks are initiated at defect locations within the plastically deformed material. The micro-cracks usually occur at the contact surface or between that and the depth of maximum Hertzian shear stress. When numerous micro-cracks initiate at the surface the phase turns to surface distress.

In the final phase, a macroscopic crack driven by high Hertz shear stresses is initiated from one of the micro-cracks. This constitutes the last failure phase: crack growth. From a normal initiation point above the maximum Hertz shear stress, it travels downward into the material, at an acute angle with respect to the direction of contact travel, until it reaches the depth corresponding to the maximum (unidirectional) Hertzian shear stress.

The material volume undermined by the crack spalls out forming a crater (exfoliates), along a crack running upward at an angle of about 30º to the surface and parallel to the 30º deformation bands, to form a crater bounded by fracture surfaces [18].

In bearing and gear applications the macro-crack branches upward to the free surface to create a spall, bounded by fracture surfaces [18].

3.2. Characterization of the failure processes

Next, a synthesis of the characteristics of the types of damages referred is presented, including its causes and effects according to the scales and place where fatigue starts.

3.2.1. Subsurface and surface origin spalling

Spalling life is inherently statistical. This is the case presumably because defect severity and location, two determining variables of life, are statistically distributed among macroscopically identical contact components. [3]

According to the location of the initiating defect, two major classes of spalling failure are distinguished:

- Subsurface origin spalling: occurs from defects in the bulk material subject to the Hertzian cyclic stress field. The defects most likely to produce this type of spall are located near (but generally above) the depth of maximum alternating Hertz shear stress. In this type of spalling, Hertz shear stress level, material matrix fatigue resistance, defect severity, and defect location are the governing variables determining life to spalling.
- Surface origin spalling: occurs from defects in the immediate subsurface material subject to asperity-scale cyclic stress fields. These defects fall into two main classes: pre-existing defects and surface distress. It can be argued that all defects in the former class initiate spalling by first causing localized surface distress.

  In this type of spalling, factors promoting it, in addition to severe levels of the variables enumerated under subsurface origin spalling are: surface distress and high values of tractive stresses in the contact interface. Residual stresses in the subsurface layers may promote or retard spalling, depending on whether it is tensile or compressive.

3.2.1.1. Appearance of spalling

Spalls show a highly distinctive appearance. A fully developed spall (with diameter much larger that its depth) shows a distinct bottom and walls. The walls may intersect the surface at a steep or shallow angle. The spall bottom consists of a series of crack surfaces. As a whole, it parallels the contact surface at roughly the depth of the maximum unidirectional shear stress in Hertzian contact. Spalls substantially larger in area than their depth often show serrations in the spall bottom, which run transverse to the rolling direction. They are formed one at the time as a crack penetrates downward in the direction of the contact travel at an angle of about 30^o until it reaches the spall bottom level, and then the undermined piece breaks away along a plane rising in the direction of contact travel at an angle 30^o towards the surface. A subsequent crack forms further down in the rolling direction, proceeds down into the material and forms the next step when the undermined material breaks away. As this description implies, spalls tend to propagate in the direction of contact travel.

Subsurface origin spall

Any of the distinguishing features indicate a subsurface origin spall:

- A steep entrance wall (inclined by more than 45^o angle to the contact

surface).

- An originating subsurface defect may remain visible after the spall has formed or may be lost with the spalled material.

Surface origin spall

The distinguishing features of a surface origin spall are in the entrance zone of the spall, which may show any of the following features:

- A shallow-angle entry wall (initiated at less than 30º angle to the contact surface).
- An arrowhead configuration formed by two shallow cracks emanating from a common point defect.
- The presence of a visible surface defect (nick, dent, furrow, pit) just upstream of the spall initiation, often with a narrow ledge of unspalled surface between defect and spall.

3.2.1.2. Causes of spalling

There are some common causes for spalling fatigue, such as:

- Material (Matrix) deficiencies: *Low hardness* materials possess a very low spalling fatigue resistance. In steels, spalling resistance increases with hardness up to quite high values. In bearings only high hardness parts have exhibited satisfactory fatigue life under significant contact loads.
- Matrix inhomogeneity: In steels, soft structural constituents reduce spalling resistance.
- Contact load: Any operating condition bringing about higher than designed Hertz pressures on any part of the contact drastically reduces spalling life, unless the load is so low as to be below the fatigue life.
- Contact temperature: Excessively high operating temperature, at which the material's working hardness is substantially less than its room-temperature hardness, reduces spalling resistance.
- Number of stress cycles: spalling is a damage accumulation failure mode so that its cumulative probability increases with the number of stress cycles. This does not imply that the hazard of spalling within a fixed incremental number of cycles necessarily increases with total accumulated life.

Causes of subsurface origin spalls

- Inclusions: Non-metallic inclusions in steel and foreign metallic inclusions in powder metals can be spall-originating defects.
- Porosity: Voids in the matrix, are spall initiating defects just as inclusions are.

Causes of subsurface origin spalls

- Pre-existing defects: Nicks and dents, finishing marks, near-surface structural damage to the matrix and pitting from manufacturing operations.
- Surface distress: produces a large population of asperity-scale micro-cracks and micro-spalls, which act as precursors to macroscopic spalling.
- Tractive stress in contact interface: high traction may result from high sliding speed, low ratio of EHD film thickness to surface roughness height, or ineffective boundary lubrication. High traction is believed to exert a direct effect on surface-origin spalling. It is observed that the copious micro-cracking characterizing advanced surface distress does not automatically lead to massive spalling as it would if all micro-cracks propagated to produce macroscopic cracks and spalls. It is conjectured that the micro-cracks fail to propagate in depth when there is a quiescent layer of material between the micro-cracks and the high shear stresses prevail to drive the cracks. This is the case when interface traction is low. For higher interface traction, this quiescent layer is increasingly invaded by shear stresses capable of driving cracks and initiating spalling.
- Hydraulic effect on crack propagation: It is observed when collinear sliding accompanies rolling that the trailing surface, in which the sliding direction is forward (the same as the directions of contact travel) spalls first. Cracks angle downward in the direction of contact travel in this surface. The leading surface, on which sliding occurs in the opposite direction spalls later, to a lesser degree, or not at all. When cracks form in this surface, they angle downward away from the direction of contact travel.

  A persistent hypothesis to explain selectivity in the direction and frequency of cracking is that the presence of a liquid lubricant is required for crack propagation in contact fatigue. The most common form of this hypothesis, attributed to S. Way [7], postulates that in a crack extending downward in the direction of contact travel and filled with oil, the approaching contact reaches the surface opening, pinches the crack closed, traps the lubricant and puts it under hydraulic pressure which extends the crack tip. A crack extending downward opposite the direction of contact travel is squeezed empty of oil by the contact and does not propagate.

Arguments can be advanced to challenge the extensively quoted Way hypothesis. One objection is that it cannot apply to subsurface-origin cracks. Yet their frequency and direction depends upon sliding direction in the same way as for surface-origin cracks. More detailed arguments are based on recent fracture-mechanical analyses of crack propagation in the presence of oil [19], which do not appear to support the Way model.

It appears more plausible to attribute the effect of sliding direction on spalling to the asymmetry in the elastic-plastic stress condition existing under any traveling Hertzian contact. The asymmetry of deformation bands with respect to the direction of contact travel is direct evidence of this stress/strain asymmetry. Elastic-plastic analyses of the stress condition during crack propagation are under-way, for sliding either with, or against the direction of contact travel [20], but are not yet conclusive.

3.2.1.3. Effects of spalling

Spalling failure is of differing severity depending on design, material and application of the contact component. Generally, spalling directly leads to non-functionality, because it is progressive (at times propagating very rapidly after initiation) and often leads to fracture or seizure.

In small or high-speed, through hardened steel rolling bearings, and in bearings with poor lubrication, spalling is a disabling failure unless arrested by a relaxation of operating conditions. It propagates rapidly and can lead to bulk fracture and loss of functionality. Accuracy of shaft support is often compromised. Vibration and noise from a spalled component is often unacceptable. Separators of spalled small bearings may fracture and cause seizure.

In large, slow speed roller bearings with good lubrication, especially if parts are case hardened, spalling may propagate slowly and considerable service life may be obtained after spall initiation. However, catastrophic spall progression and failure is also observed in such components.

In rolling bearings, spalls do not heal. They are arrested spontaneously only if they have occurred in an area which subsequently becomes unloaded (a protruding “lip” near an edge, which wears off; a portion of the rolling contact which, due to realignment of the machine element is no longer under load).

In gears, progression of spalls to functional failure is common. However, spalling (“pitting”) conditions have been identified which do not lead to failure, specifically:

- Fine spalling ("pitting") on a new gear, due to rough as-finished surfaces, which, after running-in, become smoother and sustain further cycling without more severe spalling.
- Spalling in worm gears, which remain functional even after considerable spalling.

3.2.1.4. Spalling craters

Depending on the initiation site the craters can be divided into surface and sub-surface spalls. The surface spall grows from an entrance crack in the surface in the same direction as the movement of the rolling contact point. Figure 1.12 contains examples of surface initiated spalls. Characteristic features are according to Tallian [3] that:

- The entrance wall has a shallow angle to the contact surface, less than 30^{o}. In [7] angles between 20 and 24^{o} where measured.
- Seen from above the spall grows in a V-shape from the starting point. The side walls that grow in straight lines often meet the surface at nearly perpendicular angles.
- The exit wall, also from above, has a semi-circular profile with centre at the entrance point. It meets the surface at a nearly perpendicular angle.
- The overall appearance with a starting point, V-shaped expansion and radial exit gives the crater a shape, which is often described as a "sea shell". The "sea shell" is always oriented with the head against the rolling direction. Although this form is common it is by no mean typical. Thus other forms such as circular, long and thin, etc. do exist just as often.
- Just upstream of the initiation point a surface defect (nick, dent, furrow or pit) is often noticeable, often with a narrow bridge of unspalled surface between the defect and the spall.
- The spall bottom consists of a series of crack surfaces, often with small "wing" cracks, which have not continued to propagate, see Figure 1.12b. The bottom runs parallel to the surface roughly at the depth of the maximum Hertzian shear stress.
- The length and width of the spall is often substantially larger than the depth.
- For gears the spalls often appear next to each other in bands crossing the gear tooth flank with starting points just below the pitch circle.

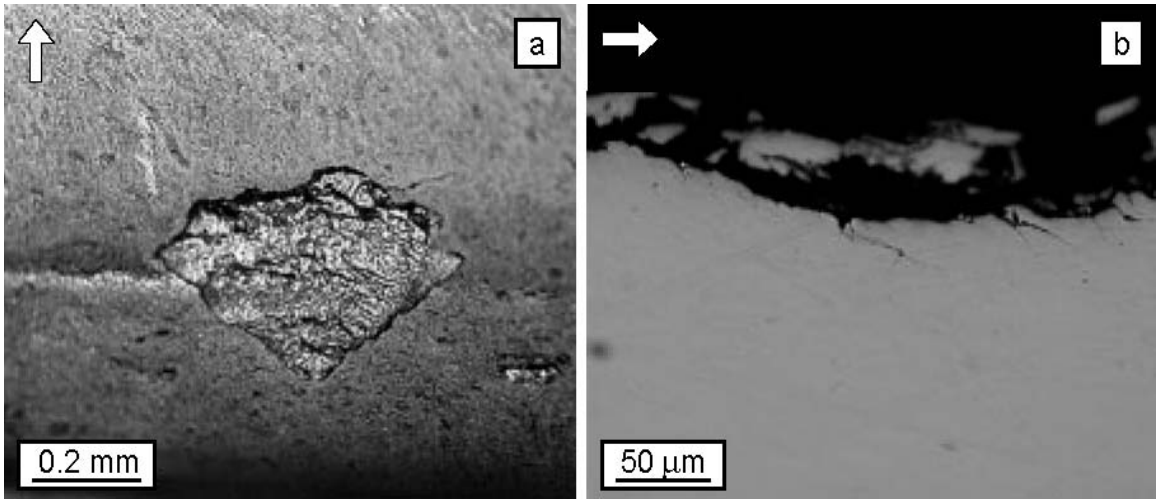


Figure 1.12 – Spalling damage in the surface of a pinion gear tooth. a) Top view of spall. b) Cut along the symmetry line of the spall. Arrows indicate the contact movements [12].

The sub-surface spall has many similarities with the surface initiated spall, in particular regarding size, characteristics of the bottom, sides and exit wall. However, it distinguishes from the surface initiated spall by the following features [3]:

- The sub-surface spall has a steep entrance wall, often with an angle above 45^{o} to the contact surface.
- The originating defect may remain visible on the bottom or may have been lost with the spalled material.
- The typical overall "sea shell" form is less common.

3.2.2. Surface distress

Surface distress is defined as micro-scale spalling fatigue [3]. It is a failure of rolling contact metal surfaces by the formation of 1) a glazed (burnished) surface areas, 2) asperity scale microcracks and 3) asperity scale micro-spall craters, in that sequence. It is the result of predominantly normal contact stresses acting in asperity dimensions.

Surface distress is widely used to designate this failure mode in rolling bearing technology. Another designation for the early, plastic flow stage of surface distress failure is *glazing*.

Surface distress has been reported almost exclusively on steel rolling contact surfaces.

Two current theories describing surface distress in metal are by Tallian [14] and Suh [21]. Both define the failure process as fatigue in asperity dimensions but differ in details of the failure mechanism.

When two surfaces are in predominant rolling contact under load and with lubrication, their asperities approach each other in a direction close to the surface normal. Depending on the average thickness and rheology of the (liquid or solid) lubricant layer between the surfaces, stresses substantially exceeding the macroscopic local Hertzian stress may develop in asperity encounters.

If the lubricant is an EHD film and its thickness is 3 or more times the composite RMS surface roughness height, then the extra asperity stresses are statistically negligible. As the EHD film becomes thinner, or if only a boundary film is present, then the asperity stresses (which can be modeled as micro-scale Hertzian stress fields) become dominant in an increasing fraction of the asperity encounters.

For asperity encounters in which the alternating shear stresses exceed a limit that depends on material properties, the following failure sequence is initiated:

1. *Asperity-scale plastic flow*: Both cumulative and alternating plastic flow occurs in the asperities.
    a. The *cumulative* plastic deformation occurs early in cycling and leads to flattening of asperity tips, especially sharp "feather edges" that often decorate the crests of asperities in ground surfaces. The appearance of the surface will become progressively more burnished (*glazed*), as the topography left from surface finishing is gradually leveled and only smooth low-frequency waviness and scratches at the bottom of deeper finishing marks remain. The surface and immediate subsurface material are heavily cold worked.
    b. The *alternating* plastic flow produces structural changes on the asperity scale which are qualitatively similar to those suffered by macroscopic Hertz contacts under fatigue stressing.
2. *Micro-cracking*: With continued cyclic stressing, the plasticity of the material is exhausted and micro-cracks form. These tend to run surface-parallel, at depths comparable to that of the high *asperity-scale* shear stresses. They may or may not be open to the surface. A heavily distressed surface is densely populated with these microcracks.
3. *Micro-spalling*: As microcracks grow and proliferate, the surface becomes undermined in asperity dimensions, and multiple microscopic spalls form.

3.2.2.1. Appearance of surface distress

Macroscopically, plastically worked and micro-cracked surfaces both appear glazed. A micro-spalled surface appears pitted. All three may be present side-by-side on the same specimen.

Glazed areas may be *generalized* on the contact surface (generalized surface distress); or, they may be *local*, in the form of *halos* around local depressions (dents, nicks) or glazed bands following furrows.

For the unaided eye the surface appears smooth but gray or "frosted". When investigated through a microscope, either parallel cracks perpendicular to the rolling direction, butterfly formed cracks or micro-scale spalls may be recognized (see Figure 1.13a). In perpendicular cuts through the frosted surface micro-cracks can be recognised, as is exemplified in Figures 1.13b and c. According to Alfredsson [12] some typical features of the micro-cracks are that:

- They are always inclined in the forward direction of the rolling contact with a shallow angle to the surface.
- If sliding is present in the contact, then the cracks are numerous if sliding is against the rolling direction but extremely few when sliding is in the same direction as the moving contact.
- Generally, they halt after 10–30 μm.
- For pure rolling the angle to the surface is between 18^{o} and 28^{o}.
- If sliding against the rolling direction is present, then the angle is larger: 41^{o}–50^{o}.

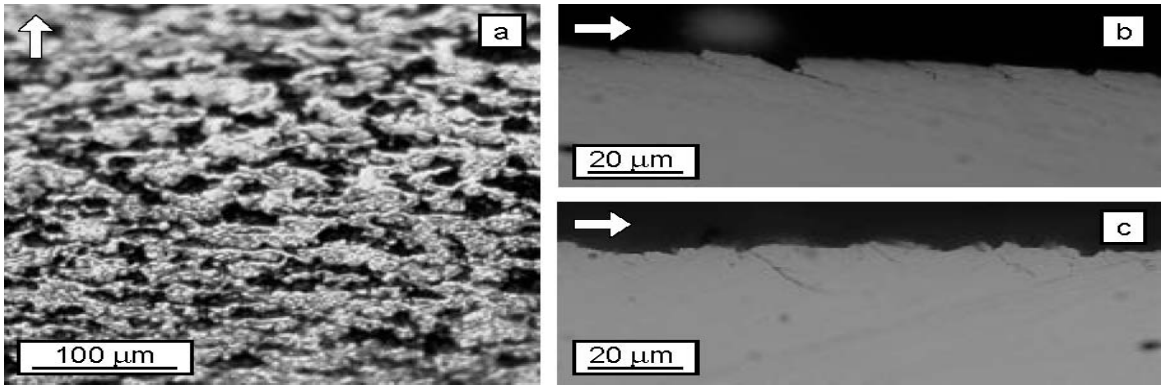


Figure 1.13 – Surface distress: a) top view, b) cut through surface distress in region with pure rolling and c) cut through surface distress in region with large negative slip. Arrows indicate the contact movements [12].

3.2.2.2. Causes of surface distress

The two dominant variables controlling surface distress are:

- *EHD film thickness to roughness ratio λ*: In the presence of a sufficiently high λ surface distress does not occur because the film prevents high micro-stresses in asperity interactions.
- *Surface microgeometry*: Both the *height* and *sharpness* of asperities (RMS roughness height and slope), influence surface distress.
    - Composite *roughness height* (RMS) enters the calculation of λ. Above a ratio of 3, *generalized* surface distress does not occur. The likelihood of surface distress increases as λ drops below this value. Local surface distress can occur where local surface features are high compared to the average, or where the film is depleted by a "sink", such as a sharp surface depression or edge.
    - Composite sharpness (RMS slope, tip curvature) of asperities influences asperity stress levels for a given value of λ. Since slope and roughness height are correlated for a given finishing method, surface distress is unlikely between very smooth, unblemished surfaces.

Other factors influential in surface distress are the *surface material fatigue resistance*, the *boundary lubricating ability of the lubricant* and the existence of *solid lubricant coatings*.

3.2.2.3. Effects of surface distress

The most hazardous consequence of surface distress is *spalling fatigue*. In the absence of pre-existing surface defects (nick, grinding burns, etc.), surface-origin spalling initiates from surface distress, through the process of asperity-scale fatigue.

The rapidity with which surface distress leads to surface-origin spalling varies greatly. The principal operating factor influencing the rate of spalling from surface distress is *tractive* interface stress. When traction is high, spalls form rapidly to appear from many initiation sites and can destroy a component in short order.

If early spalling does not terminate the operation of a component that has suffered surface distress, then surface material delamination may occur. The original surface is thereby lost over wide areas and the component becomes

unserviceable through loss of dimensional accuracy, through noise, or through secondary failures.

Gears and cams appear less sensitive to the consequences of surface distress than rolling bearings. This may be due to the greater ductility of lower-hardness gear and cam steels. However, surface distress is observed on both of these machine elements.

3.3. Influent parameters

Although already referred above for the different failure processes (spalling and surface distress) here is presented a more detailed synthesis of the parameters that influence RCF.

3.3.1. Influence of the load

In a smooth contact, load is the main parameter influencing lifespan. Contact's life can be reduced by a load increase if the Hertzian pressure is higher than the limiting pressure that leads to an infinite lifespan [9].

For a smooth contact, the maximum shear stress is located under the surface and its variation depends on the amplitude of the Hertzian pressure. For a rough contact, other shear stress maximums appear closer to surface. The amplitude of shear stresses does not depend anymore only on the load or the Hertzian pressure on the surface but also on the height of roughnesses. In the presence of significant roughnesses, even lower loads can originate fatigue damages [6].

3.3.2. Influence of the surface roughness

Surface texture is the combination of fairly short wavelength deviations of a surface from the nominal surface. Texture includes roughness, waviness, and lay, that is, all of the deviations that are shorter in wavelength than form error deviations (see Figure 1.14).

Surface roughness includes the finest (shortest wavelength) irregularities of a surface and it generally results from a particular production process or material condition.

Waviness includes the more widely spaced (longer wavelength) deviations of a surface from its nominal shape. Waviness errors are intermediate in wavelength

between roughness and form error. Note that the distinction between waviness and form error is not always made in practice, and it is not always clear how to make it.

Lay refers to the predominant direction of the surface texture. Ordinarily lay is determined by the particular production method and geometry used.

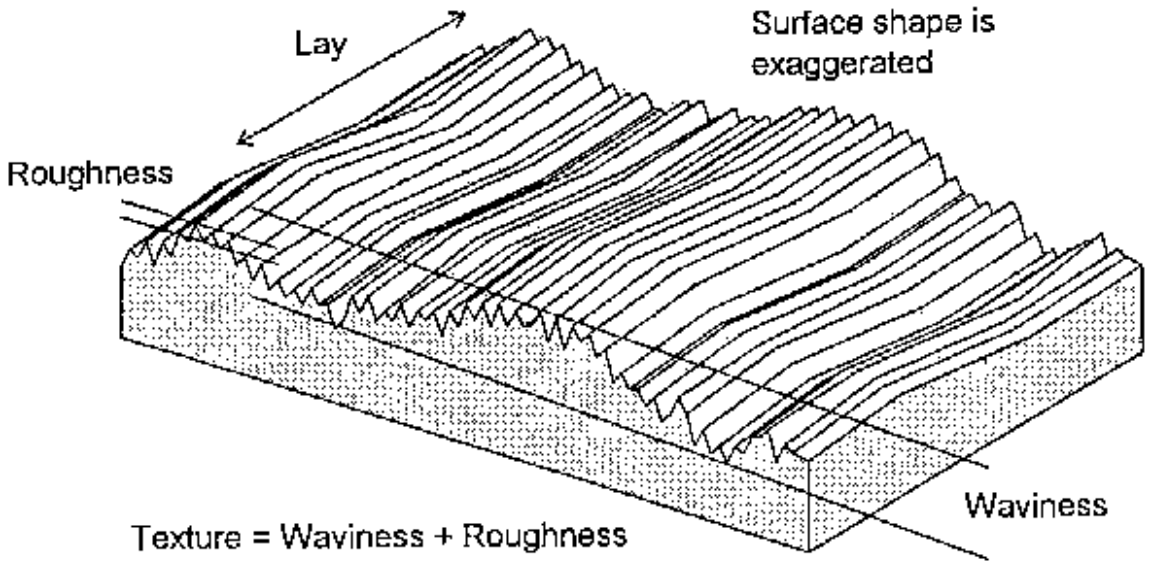


Figure 1.14 – Surface texture (waviness + roughness).

Commonly, the roughnesses of the two contacting surfaces are added to form the so-called composite root mean square roughness, *RMS* or R_q.

$$R_q = \sqrt{\frac{1}{L}\int_0^L r^2(x)dx} \tag{1.13}$$

where r is the combined deviation from the mean surface height. As it was referred in Section 2.4., the surface roughness should be investigated in combination with the lubricant film thickness by calculating the film thickness ratio λ. In Table 1.1 and Figure 1.8 the relationship between λ and surfaces damages as well as its effects on contact fatigue life were respectively presented.

Some combinations of load, slip, lubrication and surface topology may lead to wear. Fan *et al.* [22] noted that the initial contact fatigue crack under some conditions might be halted or removed by wear.

3.3.3. Influence of sliding

The influence of sliding in the lifespan of a contact is considerable. As it was already referred in Section 2.1., if sliding occurs between the contacting surfaces, tangential shear stresses at the surface will appear and the maximum shear stress will move from depth to a level close to the surface.

Sliding or slip is defined by the local relative tangential motion between the surface elements in contact:

$$s_1 = \frac{v_1 - v_2}{v_1} \tag{1.14}$$

where v_2 and v_1 are the local velocities at the contact point of the two bodies. Note that equation (1.14) gives the slip of body 1. If locally $v_2 > v_1$, then a negative slip exists on body 1.

The presence and sign of the slip will influence the number of cycles required for surface distress to appear. Negative slip shortens the number of load cycles needed for surface distress to initiate and increases the amount of surface distress for a given load level. Positive slip on the other hand counteracts the initiation of contact fatigue cracks. Pure rolling or negative slip also gives different behaviour of surface distress micro-cracks [12].

3.3.4. Influence of the material hardness

Experimental work showed that the hardness of the surfaces must be high enough to ensure the resistance of the material to surface static loads and to fatigue [6].

Tests carried out by Lorösch (see for instance Dumont [6]) on an AISI 52100 steel showed that an optimum interval of hardness led to the highest lifespans. For the same material, within the framework of rolling fatigue started on the surface, Cheng [23] showed that specimens of high hardness up to 62 HRc presented smaller cracks and a propagation speed lower when compared with those of samples of hardness of about 56.5 HRc (see Figure 1.15).

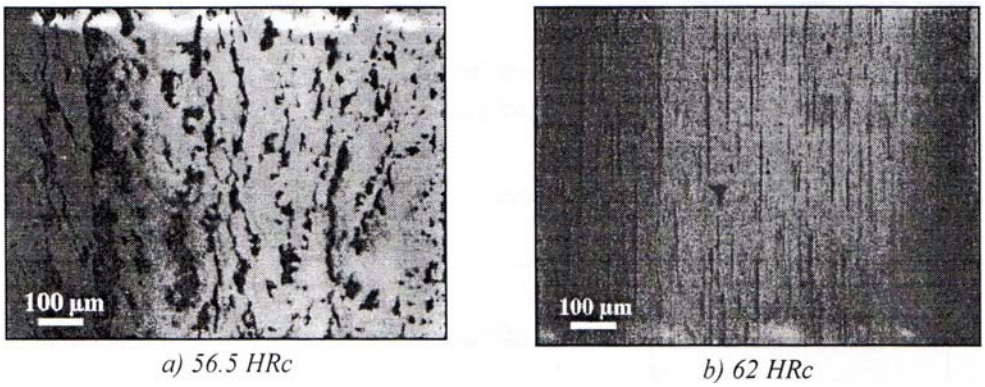


Figure 1.15 – Influence of surface hardness [23].

It should be noted that the specimens of higher hardness have higher compression residual stresses on the surface favorable to an extension of the lifespan. According to Cheng [24] two discs with 62.5 HRc and 56.6 HRc hardness presented, respectively 670 MPa and 170 MPa values of residual stresses at the surface.

In the case of mixed and boundary film lubricant regimes, a high wear rate is enhanced by a low hardness of the surfaces. Fan *et al.* [22] noted that the initial contact fatigue crack under some conditions might be halted or removed by wear under certain conditions of load, hardness and crack configuration. Wear can have a favourable effect on stopping cracking if the Hertzian pressure is not high, hardness is sufficiently high and the length of the cracks is limited.

3.3.5. Influence of material inclusions

The origin of spalling starting at a subsurface level is often due to the presence of inclusions (Figure 1.16). Their presence is harmful also due to the fact that they are located near, but generally above, the depth of the maximum Hertzian shear stress [3].

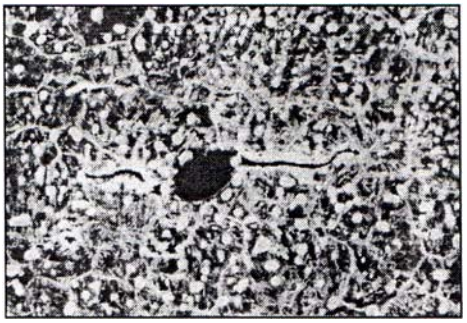


Figure 1.16 – Influence of an inclusion on crack initiation (diameter of inclusion = 5 µm) [3].

The stress field around inclusions in the area of maximum Hertzian shear stress has been investigated numerically by several researchers. For example, Clarke *et al.* [25] studied the role of the inclusions located near the surface on spalling of AISI 52100 gears teeth. Observations showed that inclusions originate damages located at the surface, particularly for aluminium inclusions with a size of 10 μm of diameter. The results of their work showed that even inclusions of low size can have an influence on the fatigue resistance of gears working under high loads and boundary regime lubrication conditions.

3.3.6. Influence of the residual surface stresses

If an unhardened steel surface was used with the Hertz contact pressures of rolling contact applications (P_0 =1.5-2.5 GPa), then severe plastic deformation would appear. Therefore, the material is case hardened, which gives increased surface hardness combined with residual compressive stresses present at the surface [12].

It should be noted that although the hardness and the residual stresses from heat treatment are results of the same process, the material hardness is only dependent on the local material state whereas the residual stress distribution is dependent on the geometry through the equilibrium of expanded and unaffected material [12].

An increasing resistance against surface fatigue with hardness was already documented by Way [7] in 1935. It was also found that spalls were smaller for a hardened surface.

3.3.7. Influence of the lubricant

Lubrication has a major influence on the lifespan of a contact. The physical properties of the lubricant: viscosity, piezoviscosity and thermoviscosity play a direct role in the oil film thickness and therefore on the lubrication regime and in the value of the film thickness ratio λ.

Cheng *et al.* [26] showed the direct influence of the lubricant's viscosity on the propagation in depth of cracks initiated at the surface. A lubricant with a higher viscosity for a given temperature will provide a much better resistance to surface damaging than another with a lower viscosity for the same temperature.

The chemical composition of lubricants, particularly the additives, play a very important role in the improvement of fatigue resistance. In a contact, many interactions take place between the lubricant and surfaces. Moreover, the

operating conditions (pressure, temperature, etc.) enhance these reactions. These chemical phenomenons are the origin of the protective film formation on the surface. In depth, they might act on the faces of the originated cracks.

3.3.8. Influence of lubricant contamination

The presence of contamination in the lubricant is one important parameter influencing the life in field applications. Contaminating particles will create plastically formed dents on the contact surfaces when rolled over by the contacting bodies. The depth and steepness of these deformations will determine the reduction in application life, which can be reduced up to 90 % [12].

The three important particle parameters: size, shape and hardness may determine the severeness of a particular contamination [12].

- Larger particles will obviously leave larger deformations on the contact surfaces. The damage is, however, dependent on particle breakthrough of the EHD film. If the contamination size is smaller than the film thickness, then the particle will pass through the contact without leaving any plastic deformation and there will be no life limiting effects.
- Compact or spherical particles are more damaging than flat particles, which, if thin enough, may pass through the contact without leaving dents.
- When passing the contact, soft particles are deformed to a thin lens shape that creates substantially less damage than spherical particles. For a hard particle on the other hand the resulting permanent dent will be deeper.

An experiment by Sayles and Macpherson (see for instance Alfredsson [12]), with large contamination particles present for the initial 30 test minutes followed by filtration to below the film thickness for the following test period showed little improvement in life compared to a test with large contamination present throughout the test. This suggests that the indent damage caused by the contamination in the initial phase of the testing is the cause of failure rather than the continued presence of contamination.

Lorösch (see for instance Alfredsson [12]) performed tests on fine surfaces subjected to comparable low loads. When full EHD lubrication film separation was introduced through increased viscosity the life was increased with, at most, a factor 4. In a second step the lubricant was thoroughly filtered to remove all contaminants larger than the film thickness. This increased component life at least 128 times. For lower viscosity and less clean lubricants, the spalling had been surface initiated, originating from noticeable surface damages. For higher viscosity and clean lubricants the spalls were sub-surface initiated, indicating the presence of another failure process.

4. Rolling contact fatigue models

Contact fatigue phenomenon is a complex and random process involving a wide number of variables. Several authors carried out important experimental and theoretical work in order to understand how all these variables influence contact fatigue. Throughout the years it was clear that two classes of models were established:

- The first ones are related with the characterization of the endurance of a material and are the so called *"Engineering Models"*. They confine themselves to variables for which extensive experimental (life test) data and field experience exists. Failure process descriptions are simplified. Many models are proprietary: only their results are published; assumptions may not be explicitly stated and the field of application may be limited [28]. In these models there is no distinction between initiation and propagation phases and the final rupture which is regarded as the end of the lifetime of a part. The work of Lundberg and Palmgren [8] constitutes the base of this first axis.
- The second type of models is the *"Research Models"* which confine themselves to a limited segment of the failure process (e.g., crack propagation only), to which the research is directed; they are often based on limited experimental background and may not attempt to correlate quantitatively with field experience [27]. They are based on the analysis of the mechanisms of cracking by fracture mechanics and dislocations theory approaches and a distinction is made between the phases of crack initiation and propagation.

4.1. Engineering models

The broad concept common to the engineering models is: The contact material is a matrix with defects. The contact applies cyclic stresses to a material volume. As stress cycling proceeds, irreversible changes ("fatigue damages") accumulate near the defects and, at the "most severe" defect, a spall forms. Defect severity, density and as a result, spalling life, are statistically distributed. Most engineering models assume linear damage accumulation (Palmgren-Miner hypothesis) so that the order in time-varying loading is irrelevant. A two-parameter Weibull distribution is commonly assumed [27].

4.1.1. Lunberg-Palmgren baseline

The baseline for engineering models remains the Lundberg-Palmgren (LP) model [8]. Based on the work of Weibull, they postulated that fatigue cracking commences at material weak points below the rolling contact surfaces. The amount and types of weak points in bearing raceway and rolling element materials are functions of material chemical composition, metallurgical structure and homogeneity. The greater the volume of material stressed, the greater also, according to Weibull, is the risk of fatigue failure. Based on the Weibull theory, LP [8] developed the following equation which underpins all their subsequent developments; this basic equation relates fatigue life in number of stress cycles endured to the first spall, applied stress and probability of survival:

$$\ln\left(\frac{1}{S}\right) \propto \frac{N^e \tau_0^c V}{z_0^h} \tag{1.15}$$

where

S = probability of survival

N = number of stress cycles endured

τ_0 = maximum orthogonal shear stress under the contact surface

z_0 = depth from the surface to the location of τ_0

V = volume of material stressed

The exponent e is called the Weibull slope or shape parameter. According to LP [8], e = 10/9 for point contact bearings and 9/8 for line contact bearings. The exponents c and h are determined empirically; they are dependent on the Weibull slope value e.

The principal equation resulting from the LP [8] effort, derived from equation (1.15), relates the bearing fatigue life to the applied load:

$$L_{10} = \left(\frac{C}{P}\right)^p \tag{1.16}$$

where

L_{10} = 90 per cent reliability bearing fatigue life in 10^6 rev.

C = basic dynamic capacity.

P = equivalent load.

p = load–life exponent.

Equation (1.16) considers that not all the bearings subjected to applied load P will survive to fatigue life L. Rather only 10 per cent of the bearings are expected to survive L. L, also designated L_{10}, is called the *rating life*.

To take into account the evolution of technology, the LP models were developed by introducing correction factors, *life factors*, a_1, a_2, and a_3. The following equation was proposed to predict bearing fatigue endurance better:

$$L_{na} = a_1 a_2 a_3 \left(\frac{C}{P}\right)^p \tag{1.17}$$

where

a_1 = life adjustment factor for the selected reliability level.

a_2 = life adjustment factor for the material.

a_3 = life adjustment factor for the operating conditions.

L_{na} = adjusted bearings fatigue life in 10^6 rev.

4.1.2. The Ioannides-Harris theory

Unfortunately, the modified life equations are not able to predict the bearing fatigue life for any given bearing over a complete range of operation. For example, at very small loads, such as those experienced by electric motor ball bearings, in the absence of lubricant contamination, incorrect or misaligned mounting and lubrication interruptions, the occurrence of RCF is non-existent. Bearings lives predicted by equation (1.17) are always finite. In the 1980s it was demonstrated that bearings manufactured from very clean vacuum-degassed AISI 52100 steel did not fail from fatigue, even when loaded to maximum Hertz stresses far in excess of normal operation. Hence, it appeared that, similar to structural fatigue, RCF is also influenced by fatigue limit stress. This means that, if a component in rolling contact does not experience stresses in excess of the fatigue limit for its material, fatigue failure will not occur [4].

In their 1985 publication, Ioannides and Harris (IH) [9] included the effects of fatigue limit stress. For a volume element ΔV_i sufficiently large to contain a big amount of defects, the probability of survival ΔS_i after a number of cycles N is given by the following equation:

$$\ln\left(\frac{1}{\Delta S_i}\right) \propto \frac{N^e \left(\sigma_i - \sigma_u\right)^c \Delta V_i}{z_i^h} \tag{1.18}$$

where

σ_i = stress at volume element i.

σ_u = fatigue limit stress.

e, c, h = material parameters.

z_i = distance below the surface to volume element i.

Compared with the LP equation (1.15), the IH equation has the following differences [4]:

1. The use of σ as the failure-initiating stress implies that maximum orthogonal shear stress τ_0 does not have to be used as the failure-initiating stress. Rather, some other stress, e.g. the von Mises stress σ_{VM}, may be used.
2. Equation (1.18) contains a fatigue limit σ_u, similar to fatigue life equations describing structural fatigue. This implies that unless, at a given location i in the material $\sigma_i > \sigma_u$, fatigue failure of the volume element ΔV_i will not occur, and the probability of survival, ΔS_i, of ΔV_i under stress σ_i is 100 per cent of 1.
3. When τ_0 is selected as the failure-initiating stress and $\sigma_u = 0$, equation (1.18) is identical to equation (1.15). This implies that the LP equation is a special case of the more general IH equation.

4.1.3. Tallian rolling bearing model

Tallian also developed a model based on the LP model [28]. The main concepts of Tallian's model are [27]:

1. Spalling life at a defect is related to defect severity and to the plastic strain created at the defect by the applied alternating shear stress. Spalling life of a component is derived from the statistical description of the defects in it.
2. Spalls may originate from surface or subsurface defects.

Tallian model uses the maximum shear stress as the plasticity criterion and the Mason and Coffin propagation law. The model's equation is:

$$C\frac{dA}{dN} = \left(\frac{\gamma}{D}\right)^{\zeta} \tag{1.19}$$

where

C = constant.

A = size of the crack.

N = number of cycles.

γ = plastic deformation.

D = ductility of the material.

ξ = Coffin-Manson law exponent.

Tallian considers that the surface and the sub-surface are favorable to have two independent fatigue processes. The probability of survival for the two failure modes is deduced from the probabilities of survival in the sub-surface and in the surface by the equation:

$$\ln S = \ln S_{SC} + \ln S_S \tag{1.20}$$

where

S_{sc} = probability of survival in the sub-surface.

S_s = probability of survival in the surface.

Later Tallian developed another model that he called "the new model" [28]. According to him, the new model is intended to be suitable as a "least common denominator" for life prediction, by incorporating the most commonly accepted failure mechanisms, functional relationships and variables into an explicitly derived and readily calculated closed-form life prediction method. After insertion of experimentally supported parameter values, the model is intended to serve as a practical life calculation tool.

4.1.4. Other models

Besides the referred models others were developed based on the LP baseline model. Among those are the ones listed below:

- Schlicht-FAG bearing model;
- Timken Tapper roller bearing model;
- Townsed *et. al* hardened steel gear model;
- AGMA/ ANSI Standard gear rating model.

A more detailed explanation of each one can be found for instance in Tallian's review of published models [27].

More recently, another method, the Stress-Life method, based on the IH one was developed by Harris *et al.* [4].

4.2. Research models

Most reviewed research models are fracture-mechanical expressions of crack formation. All such models describe crack initiation and propagation.

4.2.1. Fatigue cracking mechanisms

The fatigue cracking mechanisms generally rely on two approaches:

1. linear elastic fracture mechanics;
2. dislocations theory;

4.2.1.1. Linear elastic fracture mechanics

Fracture mechanics are used to model the macro-crack propagation phase. If the relative size of the plastic zone at the crack-tip is small compared to the structure, then Linear Elastic Fracture Mechanics (LEFM) may be used. Hereby, the critical crack length, growth rate and direction of growth may be estimated given the structure load. In general a crack tip can be subjected to three types of loads, or modes (see Figure 1.17). In the case of spalling the crack tip loading is a combination of mode I and II, symmetric and anti-symmetric in the crack plane.

For quasi static LEFM under mode I loading, crack growth may appear when the stress intensity factor (SIF)

$$K_I = \lim_{x \to 0} \sigma_{yy}(x,0)\sqrt{2\pi x} \tag{1.21}$$

reaches the critical value K_{Ic}. Corresponding criteria can be used for mode II, III or even mixed-mode loading. In the spalling process this criteria will determine when the final branching of the macro-crack towards the surface will happen.

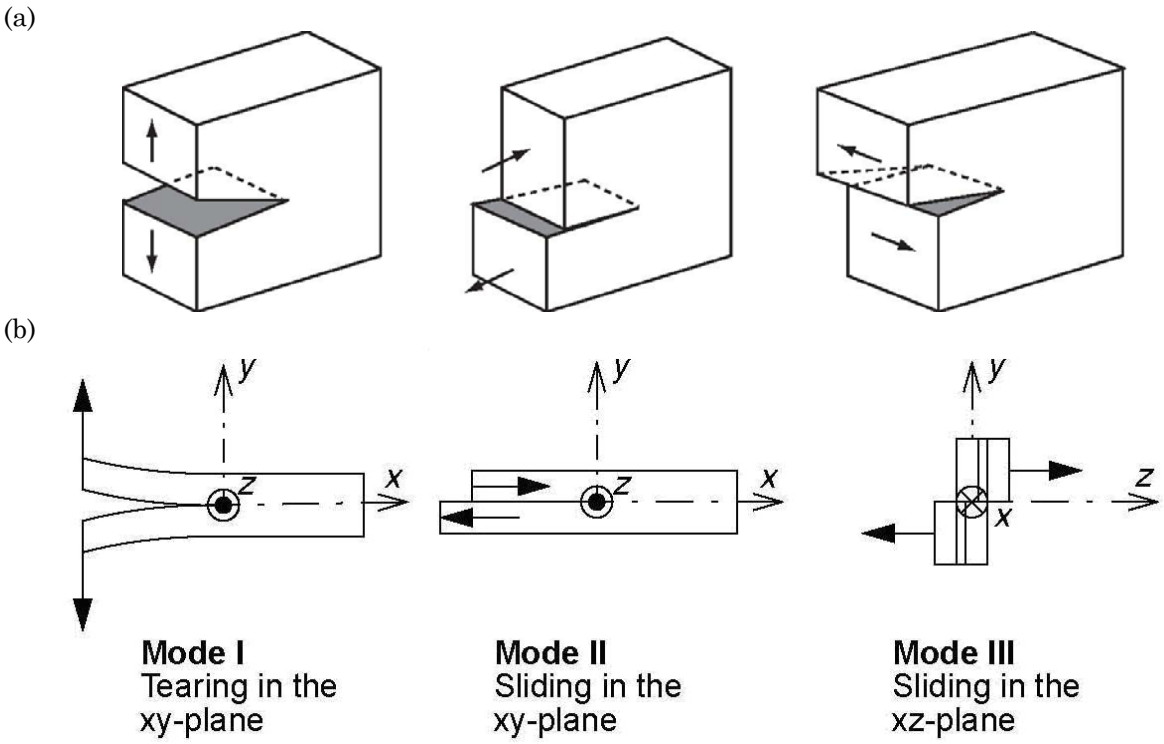


Figure 1.17 – Fracture mechanic modes in three-dimensions (a) and two-dimensions (b).

Prior to final crack branching fatigue crack propagation determines the spalling process. In mode I, crack propagation is driven by the SIF range, ΔK_I. If ΔK_I is above a threshold value, ΔK_{th}, then the empirically found Paris law

$$\frac{da}{dN} = C(\Delta K_I)^n \tag{1.22}$$

where da/dN is the crack growth rate, C and n are material parameters, is valid.

Combined mode I and II fatigue loading, as exist during the spalling macro-crack propagation phase, has been reviewed by Bold *et al.* [29] and by Qian *et al.* [30]. These reviews indicate that the possibility to propagate a crack in another direction than mode I is dependent on material, load magnitude and loading sequence. It requires a ductile material subjected to large loads under symmetric load sequence. Thus, for the current situation with brittle steel subjected to mode I and non-symmetric mode II load, other directions of crack propagation than that of mode I is unlikely.

Gao *et al.* (cited by Alfredsson in [12]) performed mixed-mode I and II fatigue propagation experiments on steel. When plotting the normalized SIF ranges

$\Delta K_{II}/\Delta K_{th}$ against $\Delta K_I/\Delta K_{th}$, three different areas of crack growth were found (limited by two curves in Figure 1.18). Below the first curve no growth took place, between the two curves growth was strongly influenced by crack face friction which resulted in crack arrest, and finally above the second curve the crack turned to a growth direction with predominantly mode I loading. Thus, the effect of a mode II field is two-fold: Firstly, the crack will turn towards the pure mode I direction and secondly, the mode II field reduces the mode I threshold, ΔK_{th}, and therefore promote crack propagation.

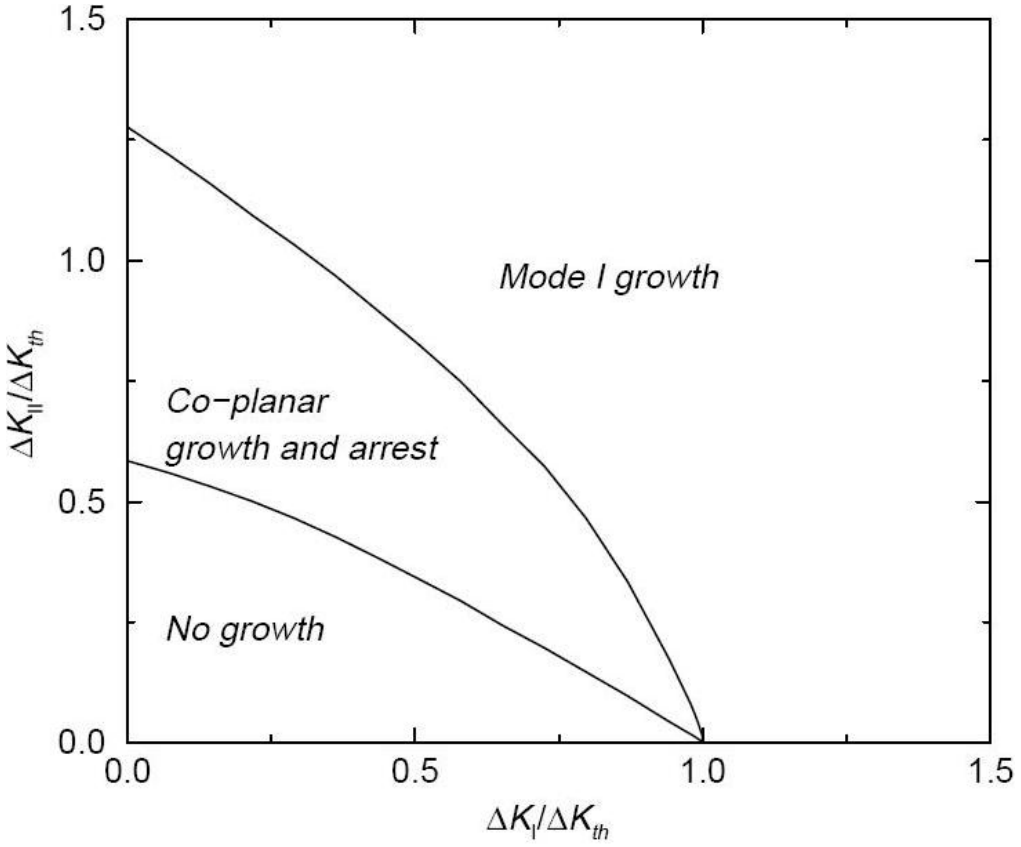


Figure 1.18 – Normalized SIF ranges in combined mode I and II [12].

4.2.1.2. Dislocations theory

This theory, based on micro-structural considerations, treats the initiation and propagation phases of fatigue cracks.

The initiation occurs when there is a movement of dislocations along the slipbands, along grain boundaries or at the interface of inclusions. Cracking appears when the density of dislocations reaches a critical size. The base

hypothesis is that two dislocations corners of opposite signs can create a crack if the distance between them is sufficiently weak.

The propagation is modeled by the emission of dislocations located in the plastic zone of the crack tip.

The cracking modes are associated with the emission modes of the dislocations corners in the case of modes I and II or of screw dislocations in the case of mode III. Under certain loading conditions, the theory of dislocations allows quantitatively to approach the propagation velocity of the cracks as well as the deviation of the cracks compared to their orientation at the origin.

4.2.2. Crack initiation models

Mura *et al.* [31] created a model for crack initiation based on the dislocations theory.

The common topic between the works of different authors using the dislocations theory is the concept of dislocations pilling along permanent slipbands. The alternate movement of plastic flow in the slipbands interior and the load movement are modeled by dislocations of opposite signs moving on two close layers. Pilling is related with the extrusion phenomenon and with the intrusion of dislocations dipole along one cycle. The number of cycles until initiation of a micro-crack depends on the site: grain, grain joint, and interface with inclusion (Figure 1.19). It is reached when the deformation energy due to the dislocations accumulated after N cycles reaches a critical value.

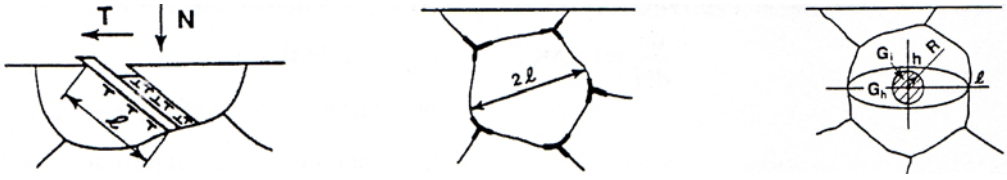


Figure 1.19 – Micro-crack initiation sites [32].

In the case of RCF, Zhou [32] highlighted the transformations that occur in the matrix close to surface (Figure 1.20).

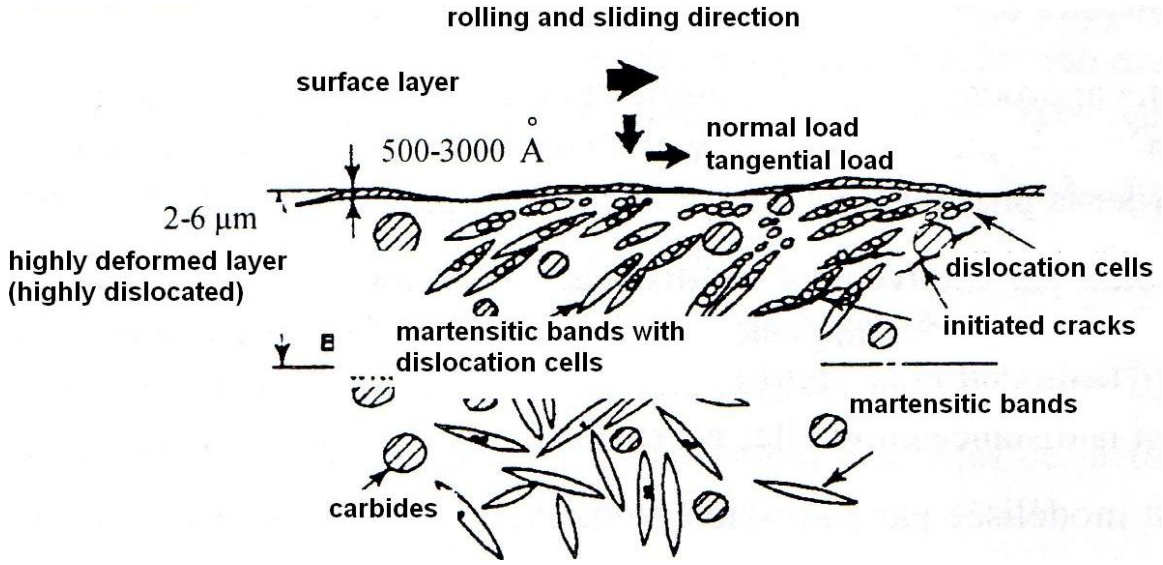


Figure 1.20 – Steel structure with dislocations cells [32].

At the subsurface level, where the cyclic shear stress is higher, the emission of dislocations at the initiation origin is induced by the incompatibility of deformation between matrix and inclusion. The existence of dislocations is therefore necessary to the interface inclusion-matrix.

If the dislocations rate is higher than a critical value, the initiation of cracks can take place in the preferential direction that corresponds to the maximum shear stress τ_{max}.

Near the surface or at the surface, in the case of mixed and boundary lubrication, shear stresses due to the asperities become prevalent. The observations show the formation of deformed layers deformed accompanied with plastic deformation, dislocation cells and microscopic cracks close to the surface.

The experimental work conducted by Cheng *et al.* [24] on the initiation of micro-cracks induced in a rough contact under pure rolling conditions show that the initiation probably takes place at the surface level in fracture mode I under the effect of tensile stresses and also in mode II, under the effect of high shear stresses.

Concerning the study of surface crack initiation, Yamashita *et al.* [33] developed a model based on the mechanisms modeled by Mura *et al.* [31], in which the influence of asperities in a Hertzian contact is considered. They highlighted the development of cracks by modes I and II. Mode I appear due to the existence of tensile surface stresses towards the edges of the contact on a very thin zone.

According to the length of the microscopic crack, two modes of cracking can then be considered: 1) mode I for very short cracks and 2) mode II for longer cracks.

4.2.3. Crack propagation models

A crack propagation model should be able to explain a number of experimentally found observations, from the moment a crack initiates until it propagates and eventually ends in a spall.

Based on experiments, Way [7], with his pioneer work, pointed out that in order to produce spalling damage it is necessary that:

- oil is present,
- the oil viscosity is below a critical value and
- the surface roughness is above a critical value.
- it was also noticed that a hardened surface has higher resistance towards spalling than an unhardened surface.

Bower [19] concluded that apart from explaining the effect of oil presence, the spalling model should also explain

- the forward direction of crack propagation and
- the life dependence on presence and direction of frictional forces.

In the continued search for a rolling contact fatigue model, Alfredsson [12] also point the following questions that seek for an answer:

- why do the surface distress micro-cracks turn into the two different ranges of angles to the surface?
- under what loading condition will a micro-crack continue to propagate in the typical direction and shape of spalls?
- do the same or two different processes govern surface and sub-surface initiated spalls?

The crack propagation models developed up till now are two or three-dimensional and based on fracture mechanics approaches and on the dislocations theory.

The major developments occurred in the 80's and among those the following models played an important role.

Keer and Bryant Crack Propagation Model

In 1983, Keer *et al.* [34] present a two-dimensional fracture-mechanical crack propagation model of rolling/sliding Hertzian contacts. Traction, crack growth direction and hydraulic effects in the crack (Way hypothesis [7]) are covered. Paris' law of crack propagation is used in the following integral form:

$$N \equiv N_0 + N_c = N_0 + \frac{1}{2\beta_0}\int_{b_0}^{b}\frac{db}{(\Delta K)^m} \tag{1.23}$$

where

N_0 = crack initiation time.

β_0, m = numeric constants.

b_0 = crack length.

b = initial crack length.

ΔK = stress intensity range during a stress cycle.

ΔK depends on the momentary crack length i.e., the Palmgren-Miner linear damage accumulation hypothesis is violated.

Later, Hanson *et al.* [35], based on Keer' approach, analyzes a three-dimensional crack propagation model in a two-dimensional rolling-sliding contact. A fatigue limit stress is assumed, with a discontinuity of crack growth rate at that stress.

Zhou, Cheng and Mura

In 1989, Zhou *et al.* [32] analyses an elliptical partial-EHD contact, on both the asperity and macro-contact scales. Crack initiation life is given by:

$$N_i = \frac{A}{(\Delta\sigma - 2\sigma_k)^2 D} \tag{1.24}$$

where

A = material constant

$\Delta\sigma$ = the (von Mises) stress range

σ_k = critical value of this range below which cracks do not initiate

D = correction factor that depends on the applied stress

Crack propagation is described by a Paris type law similar to that in equation (1.23) and spalling life is the sum of these two components which are of the same order of magnitude.

Bhargava et al. Plastic Strain/Life Model

In 1990, Bhargava *et al.* [20] studies the plastic cumulative strain in a strain-hardening material under contact stress cycling. The stress vs. plastic strain curve is a "cliff" curve rising steeply from zero at a shakedown limit stress and flattening toward horizontal near twice that stress. Spalling lives are predicted from the plastic strain.

Blake and Cheng Spur Gear Spalling Model

Blake and Cheng's model (1990) (see for instance Tallian [27]) is a crack propagation model for spur gear spalling intended to form the basis of a future engineering model. Effects of contact pressure, crack initiation angle, traction coefficient, slide/roll ratio and the film thickness ratio λ are calculated for carburized test roller sets and for gear sets. Quantitative results, including load/life law and life distribution are critically dependent on assumed material constants, initial crack size and other parameters.

Bower

In 1988, Bower [19] studies the influence of crack face friction on life using a fracture-mechanical crack propagation model based on oil entering surface cracks. Bower evaluates 3 possible ways for oil to affects crack propagation:

- the oil reduces friction between crack surfaces and will thereby simplify shear movement. Thus, the crack propagates in shear mode (mode II) driven by cyclic shear stresses from the rolling contact. Bower [19] concludes that, it would explain the dependence on slip and crack angle, but the crack is likely to lock up due to the compressive stresses. It is also likely that the fatigue crack would branch into produce a mode I crack.
- the fluid may be forced into the crack by the load with pressure on the crack faces to generate mode I stress intensities at the crack tip. This would give rise to large stress intensities and rapid crack growth. The model, is however unable to account for the spalling life dependence on movement and slip direction [19].
- Fluid may be trapped in the crack and pushed towards the tip creating a hydrostatic pressure with a combination of mode I and II stresses present. This model is consistent with the spalling life dependence on movement and slip direction. Among these 3 models, this was favored by Bower [19]. However, it was concluded that the model predicts a complicated pattern of

combined mode I and II loading, which makes it difficult to forecast the direction and growth of the crack, [19].

Kaneta, Murakami and Yatsuzuka

In 1985, Kaneta *et. al* [36] developed a three-dimensional model to study the propagation of cracks using a fracture mechanics approach and the boundary element method. The first models [37] describe the propagation of cracks in linear contacts (Figure 1.21). More recently, the case of three-dimensional elliptical and circular contacts was studied (Figure 1.22) [39]. For linear models, propagation is only possible if the friction coefficient has an opposite direction to the load movement (Figure 1.23a). For a semi-elliptic crack subjected to a Hertzian pressure field (Figure 1.21), Kaneta *et al.* showed that the crack tip remains open for the three configurations shown in Figure 1.23a. In the case of Figure 1.23a3, the friction coefficient must be lower than 0.1. In these simulations, the cracks angle is 45º and the crack length is of the same order of magnitude that the contact, although the cracks angle observed are usually close to 20º or 30º.

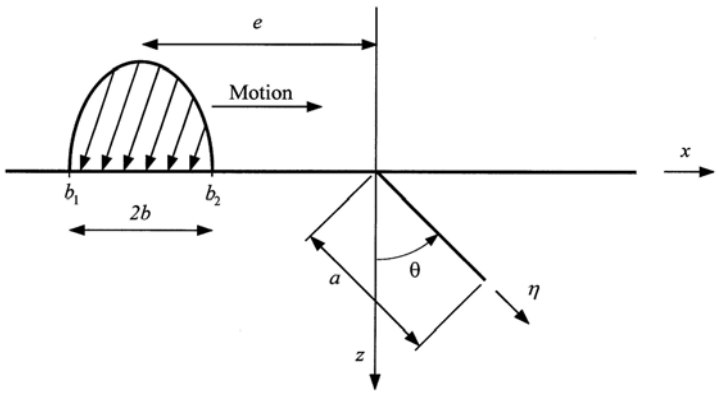


Figure 1.21– Crack propagation in 2-dimensional linear contacts [38].

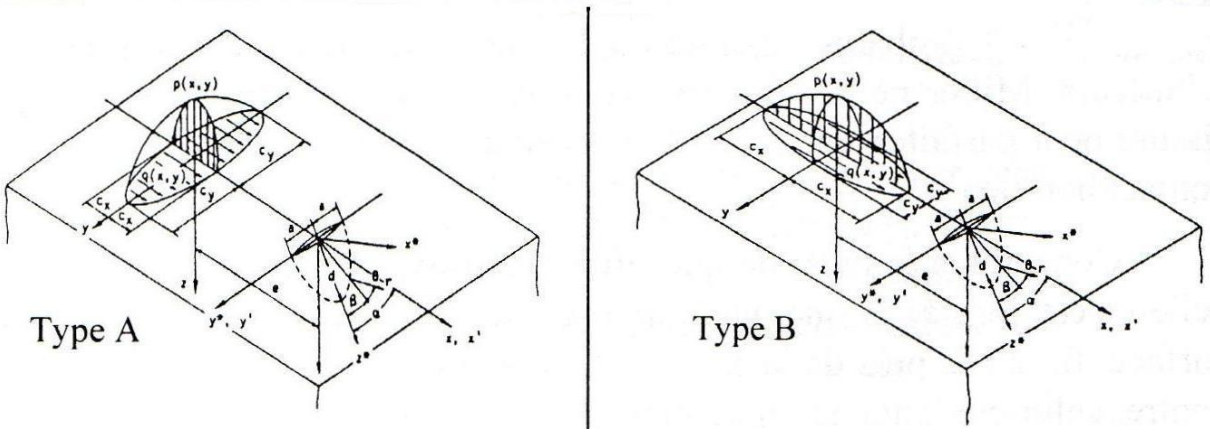


Figure 1.22 – Crack propagation in 3-dimensional elliptic contacts [39].

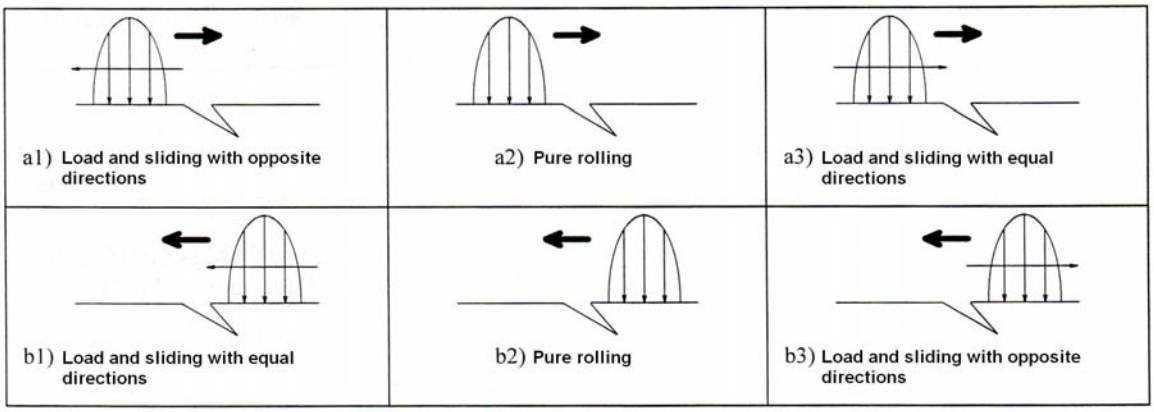


Figure 1.23 – Parameters: movement and friction directions and crack propagation angle [6].

Propagation depends not only on the possibility of lubricant penetration inside the crack but also on other factors. Furthermore, mechanisms based on oil entering cracks are unsatisfying while they do not explain how the micro-cracks appear in the first place and they can not predict the spalling crack path without introducing severe restrictions on the load pattern.

Throughout the years, several hypotheses were advanced concerning the way the lubricant acts on crack propagation. Summarizing the contribution of different authors, three modes of lubricant action on crack propagation can be referred:

- the lubricant exerts a pressure on the open crack faces and enhances propagation in mode I (Figure 1.24b) (two-dimensional analyses: Keer and Bryant [34], Bower [19], Kaneta *et al.* [36]; three-dimensional analyses: Murakami *et al.* [40]).
- the lubricant is trapped in the crack and supports a complex mode of crack propagation in modes I and II (Figure 1.24c) (2-dimensional analyses: Bower [19]; 3-dimensional analyses: Kaneta and Murakami [37,39]).
- the fluid present lubricates the crack faces, reduces friction between the crack faces due to the presence of the fluid and enhances propagation in mode II (2-dimensional analysis: Bower [19]).

It should be highlighted that the size of cracks considered in these models and in the experiments accompanying them is always high. Indeed, the initial length is always of about size of the contact. Kaneta *et al.* [37] considers that a crack has a small size if the ratio crack length/contact half-width is lower than 0.2. The influence of the lubricant was hitherto shown only for big cracks and largely open on surface. It is difficult to affirm that the lubricant intervenes at the beginning of the stage of propagation when the cracks are still very small.

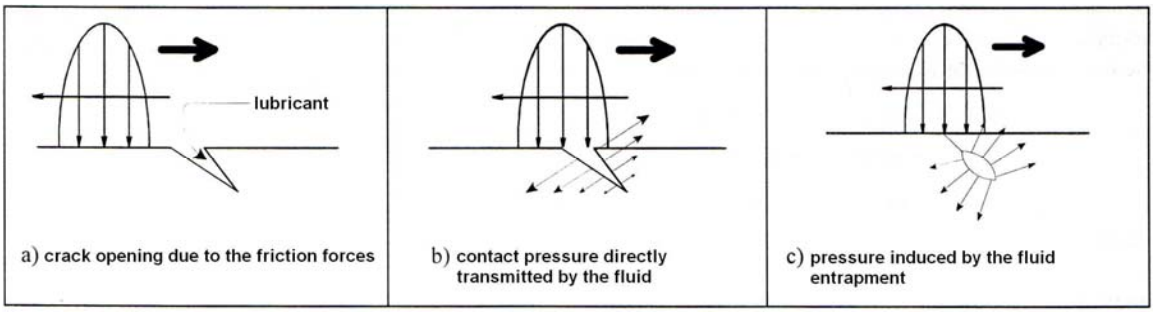


Figure 1.24 – Lubricant action in crack [37].

Micro-cracks initiation is related with the mechanism of dislocations' accumulation. It is thus a phenomenon that occurs at the material's microstructure level. Fracture cracking mode II seems to be prevalent in this stage. Even though, mode I also may appear in this initiation stage.

Concerning the propagation modes, the fracture cracking modes I and II were highlighted in cracking but not in the formation of spalling itself. It is not established yet which mode I or mode II conduit to spalling formation.

The propagation of fatigue cracks poses a problem of scale. During cycles, the size of the crack evolves by compared to that of the microstructure level and compared to that of the plastic zone formed around the crack. The mechanisms of propagation will also probably evolve. Concerning the role of the lubricant in the propagation in depth of microscopic cracks initiated at the surface, it should be noticed that the dimensions of the modeled cracks or artificially created ones for particular tests are much more significant than those dimensions of microscopic cracks observed on specimens surface that suffered spalling. Nevertheless, the role of the lubricant makes it possible to interpret the differences in damage obtained between the slower and the faster surface when sliding exists in a rolling contact.

4.3. Criteria

Along with the search for improved models, the search for criteria is also an essential stage to better evaluate the life to fatigue of contacts.

A criterion defines a border out of which damage will appear. The general formulation of a criterion is:

$$\tau + a\sigma < b \tag{1.25}$$

where

τ = shear stress in the material.

σ = tensile stress in the material.

a, b = constants obtained experimentally for a given material and number of cycles.

The fatigue criteria used are based on the Tresca and Von Mises criteria and can be divided in 2 families: 1) resistance criteria and 2) multiaxial fatigue criteria.

4.3.1. Resistance Criteria

These criteria define a limit corresponding to the appearance of the first plastic deformations in the material.

The maximum shear stress criterion, also known as Tresca's criterion, is often used to predict the yielding of ductile materials. It considers that a given point in a body is considered safe as long as the maximum shear stress τ_{max} is not higher than the yield shear stress τ_{lim} obtained from a uniaxial tensile test.

$$\left|\tau_{\max}\right| < \tau_{\lim} \tag{1.26}$$

The maximum shear stress τ_{max} is related to the difference in the two principal stresses (obtained from the Mohr's Circle).

The von Mises Criterion, also known as the maximum distortion energy criterion is often used to estimate the yield of ductile materials. It states that failure occurs when the energy of distortion reaches the same energy for yield/failure in uniaxial tension.

$$\frac{1}{2}\left[(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2\right] \le \sigma_{\lim}^2 \tag{1.27}$$

4.3.2. Multiaxial Fatigue Criteria

The Tresca and von Mises criteria are the base of several other criteria developed for multiaxial fatigue. Generally, they use the concepts of average σ_{avg} and alternate σ_{alt} components of the stress field by considering that it is alternation around the average value is the cause of material damage:

$$\sigma = \sigma_{avg} + \sigma_{alt} \tag{1.28}$$

Rolling fatigue occurs in the macroscopic elastic domain. However, at a microscopic scale, the material is not homogeneous but consists of grains, and the local stress can exceed the yield stress for certain misoriented grains. To answer this problem, Dang-Van [41] proposes a criterion of damage initiation at the microscopic level (see Figure 1.25).

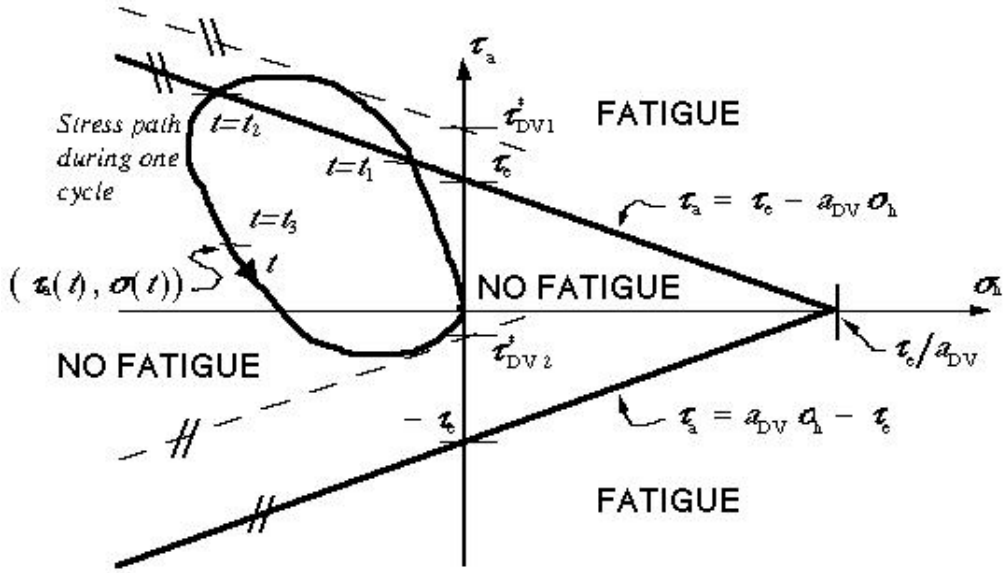


Figure 1.25 – The Dang-Van criterion [42].

Equivalent stresses (i.e. a measure of the stresses in the multiaxial state that are to be compared to test data from a uniaxial stress state) are denoted τ_{DV1} and τ_{DV2} and are used for calculation of the partial damage.

In Figure 1.25, σ_h is the hydrostatic stress (i.e. the mean value of the normal stresses in the current point) and τ_a the difference between current value of resulting shear stress in a specified shear plane and its mid value during one cycle. Damage is induced between $t = t_1$ and $t = t_2$. Material parameters are τ_e denoting the endurance limit and a_{DV} representing influence of the hydrostatic stress.

5. The use of artificial dents on the study of rolling contact fatigue

The presence of surface defects, such as dents, accelerates considerably the phenomenon of RCF, because they generate elastic pressure peaks when overruled, increasing the stress level. The maximum stresses occur closer to the surface, leading to surface initiated spalls. Based on this fact, several authors [22,23,45-51] tested different types of artificial surface dents as used them as a technique to evaluate the performance of RCF.

Fan *et al.* [22] pointed two main purposes for using artificial defects such as dents:

1. to induce fatigue cracks near the highly stressed defect site to avoid the randomness of the initiated crack locations;
2. to be used as a marker for the wear volume.

Geometry of a surface dent

As shown in Figure 1.26, the dent is analytically characterized by its depth denoted h_d, its diameter D, the diameter between peaks D' and the height of its shoulders h_s. The geometry of the dent was generated by a Rockwell C indenter.

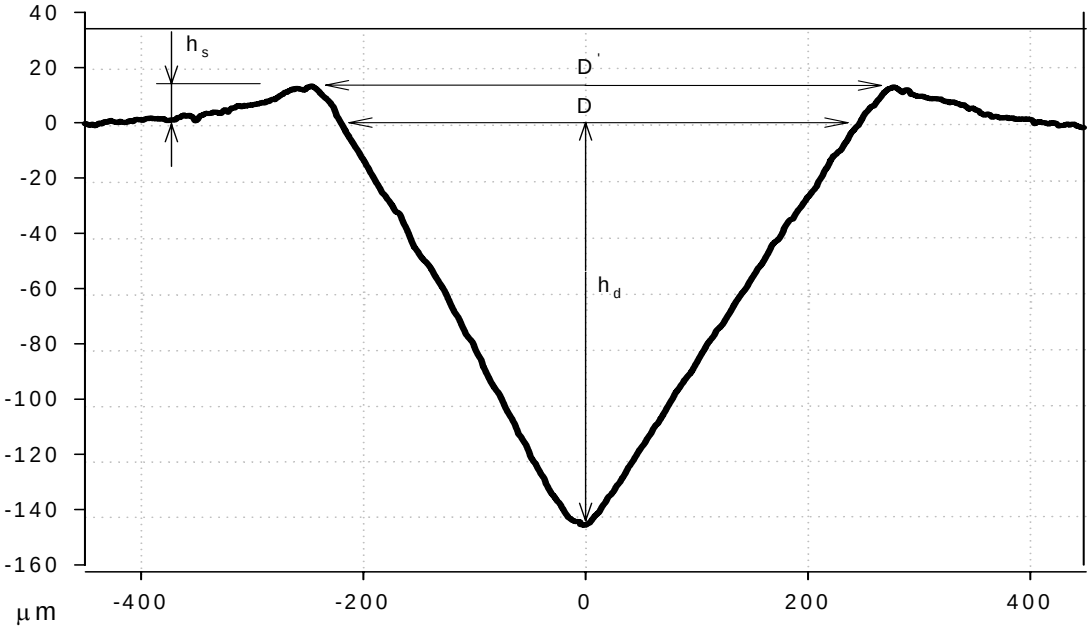


Figure 1.26 – Dent geometry.

Numerical analysis of surfaces dents indented in a surface and under rolling contact

Nélias *et al.* [51] have presented a review of the works of different authors on the numerical analysis of surface dents indented in a surface and also under rolling contact:

- Ko and Ioannides [52] were the first to perform an elastic-plastic analysis of the subsurface stress due to debris denting. In their analysis, using the finite element method (FEM), the contact problem was not solved and a given pressure distribution was considered, hence, neglecting the effect of plastic strain and subsequent hardening due to indentation.
- More recently, Sadeghi and co-workers have performed advanced numerical simulation by FEM [53-57].
- Xu and Sadeghi were the first to investigate the problem of rolling load over a surface defect [53], again without plastic strain and residual stress due to indentation.
- Later Xu *et al.* studied the indentation of thru-hardened steel by a rigid spherical tip and subsequent overrolling of the dent [54,55].
- Zhao *et al.* analyzed the effect of a dent in a line contact [56] considering the fluid entrapment in the dent pocket.
- More recently Kang *et al.* [57] determined the shape of a soft debris while traveling through the contact conjunction. Unfortunately the computational cost of such a simulation did not allow the use of a fine mesh, required to predict the plastic strain and residual stress in the vicinity of the dent. However, the macro-geometry of the dent (i.e., depth and diameter) was obtained in their study.
- Howell *et al.* [58] performed a finite element simulation of rolling contact over four differently sized spherical dents. The effect of dent size on the pressure distribution and fatigue failure location was studied for AISI 52100 bearing steel. In their analysis the material model used was adjusted to match both the stress amplitude versus strain range curve and experimental ratcheting.

Effects of a dent on the pressure distribution

Coulon *et al.* [59] studied experimentally the influence of a dent in terms pressure variation. They measured the pressure profile for a dented surface and compared it with a smooth one (Figure 1.27). As it can be seen, the presence of a dented

surface completely modifies the pressure field. From the inlet to the outlet, it was observed that [59]:

- A first pressure increase, in the area of the left dent shoulder; the pressure peak is located at the place where the lubricant film thickness is supposed to increase due to presence of the valley of the dent.
- Subsequently pressure collapses down to a very low value. This demonstrates that no cavitation occurs after the first shoulder and that a significant part of the initial lubricant volume remains entrapped inside the dent.
- Pressure starts to increase before the central part of the dent valley where the actual pressure is very close to the maximum Hertzian one; followed by a huge pressure increase leading to an intense and narrow pressure peak. The peak position corresponds to a minimum film thickness generated by the right dent shoulder.
- Finally, in the constriction region, the dented contact gives pressure values similar to the ones found with smooth surfaces.

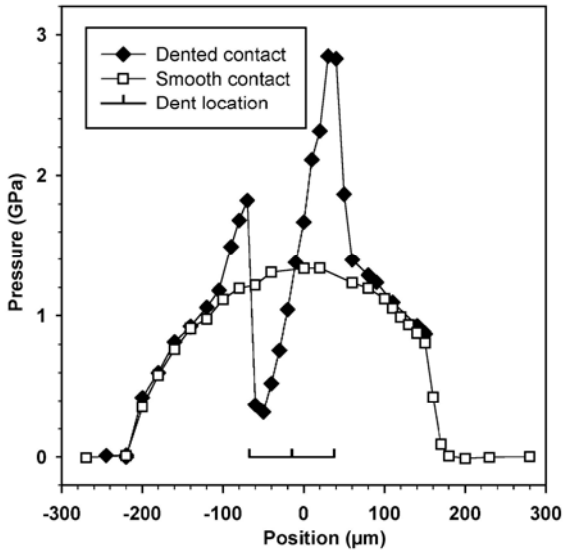


Figure 1.27 – Pressure profiles for smooth and dented contact [59].

Sites for crack initiation in a dent

The specific sites of dent initiated RCF has been demonstrated experimentally and numerically [53,60,61]. Figure 1.28 shows an artificial conical shape dent and the damaged area around it for a RCF test under pure rolling conditions. The damaged area appears in the dent shoulder and ahead the dent. These two experimental observations can be justified by the pressure modifications induced

by the presence of the dent. Firstly, the dent shoulder increases generates elastic pressure peaks when overrolled, increasing the stress level. The maximum shear stress moves closer to the surface, leading to initiated spalls. Secondly, these spalls appear ahead the dent because, as it was explained above, the second peak is higher than the first one.

Figure 1.28 – Damaged area ahead of an artificial dent.

RCF life prediction of dented surfaces

Several authors [62-65] tried to predict fatigue life of dented surfaces. Generally they use statistical modelling, for example the Lundberg-Palmgren on [8], to predict surface fatigue. However, these models do not take into account the physical aspects and only calculate stress contribution to fatigue.

Conclusions

It was shown in this chapter that significant research has been done in studying the phenomenon of RCF. A phenomenon that occurs under EHD lubrication conditions in which the contacting surfaces deform elastically while being loaded and separated by a lubricant film. The contact stresses and the geometry of the contacting bodies can be determined using the Hertzian theory and the pressure distribution and EHD film thickness in a smooth or rough contact can be determined using the formulae developed by different authors.

Extensive research also revealed that there are two scales of damages (spalling and surface distress) and three types of damage processes (surface origin spalling, sub-surface origin spalling and surface distress) originated by RCF in a EHD contact. On the other hand, several parameters (load, surface roughness, sliding, material hardness, material inclusions, residual surface stresses, lubricant and lubricant contamination) were found to influence RCF.

To study the interplay of the different parameters and predict the lifetime of a contact several models have been developed. These models can divide into two main categories: engineering models (concerning the characterization of the endurance of a material based on extensive experimental data) and research models (confined to a limited segment of the failure process, that is, crack initiation or crack propagation).

In the end of the chapter, a relatively recent technique using artificial dents was presented. This technique is based in the fact that a dent printed in a surface increases the local stress level leading to an earlier damage by RCF. A normal test with a non dented surface is much more time consuming and not feasible when results from various tests must be obtained in a short amount of time. Furthermore, the use of artificial dents is also useful as a mean to calculate the wear volume.

The subjects presented in this chapter are of great importance throughout the current work as it is summarized below:

- The formulae presented in Section 2 for the elastohydrodynamic lubrication is used to calculate the contact stresses, the geometry of the contacting bodies, the lubricant film thickness and the severity of the contact in Chapter III;
- The characterization of the types of fatigue damage presented in Section 3 is helpful to analyze the experimental results obtained in Chapter IV.
- The models presented in Section 4, and particularly the research models based in linear elastic fracture mechanics are the subject of Chapter VI in

which the SIFs of a surface-breaking crack are calculated;

- The technique presented in Section 5 using artificial dents is used in the experimental tests as described in Chapter III. In Chapter V a finite element simulation of the artificial dent used is also presented.

Chapter II

Oil and grease lubrication

1. Introduction

Lubrication plays a fundamental role in the lifetime of a contact. Basically, the function of a lubricant is to control friction and wear in a given system. These basic requirements are related to the performance of the lubricant, i.e. its influence upon friction and wear characteristics of a system. Furthermore, another important aspect concerning lubricant performance is related with its resistance to degradation in service.

There are two basic types of lubricants that can be found in the widest range of mechanical applications: oils and greases. Oils can be of two different origins, biological (mineral oils) and non-biological (synthetic oils). Greases, on the other hand, are not fundamentally different from oils. They consist of mineral or synthetic oil, but the oil is trapped in minute pockets formed by soap fibres which constitute the internal structure of the grease. It provides semi-permanent lubrication since the oil trapped in the fibrous structure is unable to flow away from the contacting surfaces.

The use of additives in both oils and greases can radically change the properties of a lubricant. Therefore, they can be essential to their overall performance.

In spite of certain limitations in performance, greases are widely used for a variety of reasons: they provide low friction, are easily confined, and have a long lubricating life at a low cost. The majority of rolling element bearings in use today (approximately 90%) are grease lubricated [66].

Given the fact that grease lubrication is widely used, it would be natural that understanding on its lubricating mechanism was already fully achieved. But, surprisingly, when comparing the current knowledge on the mechanisms of oil and grease lubrication, one concludes that grease lubrication remains a scientifically neglected area of tribology. At present only a little knowledge exists of the mechanisms of film formation, and loss, and the nature of the separating film formed. There is not satisfactory model for greases and thus it is impossible to predict performance, or service life, from simple bulk rheological properties [67].

2. Physical properties of lubricants

2.1. Oil viscosity

Oil viscosity can be defined as the measure of the oil's resistance to deformation under shear stress. This parameter plays a fundamental role in lubrication. Oils with different properties exhibit different viscosities. In addition, oil viscosity changes with temperature, shear rate and pressure and the thickness of the generated oil film is usually proportional to it. In some cases this is not true, because a more viscous oil requires more power to be sheared, leading to higher power losses an increase of heat generation, which may result in the failure of the component.

For engineering applications the oil viscosity is usually chosen to give the optimum performance at the required temperature. Knowing the temperature at which the oil is expected to operate is critical as oil viscosity is extremely temperature dependent. The viscosity of different oils varies at different rates with temperature. It can also be affected by the velocities of the operating surfaces (shear rates). The knowledge of the viscosity characteristics of a lubricant is therefore very important in the design and in the prediction of the behaviour of a lubricated mechanical system [10].

2.2. Viscosity temperature relationship

The viscosity of lubricating oils is extremely sensitive to the operating temperature. With increasing temperature the viscosity of oils falls quite rapidly. In some cases the viscosity of oil can fall by about 80% with a temperature increase of 25ºC. From the engineering viewpoint it is important to know the viscosity value at the operating temperature since it determines the lubricant film thickness separating two surfaces [10].

The oil viscosity at a specific temperature can be calculated from the available viscosity-temperature equations. The most commonly used equations are summarized in Table 2.1. Among them the most accurate is the Vogel equation. Three viscosity measurements at different temperatures for a specific oil are needed in order to determine the three constants in this equation. The oil viscosity can then be calculated at the required temperature, or the operating temperature can be calculated if the viscosity is known.

Table 2.1 – Viscosity-temperature equations (adapted from ref. [10]).

Name	Equation	Comments
Reynolds	$\eta = be^{-aT}$	Early equation; accurate only for a very limited temperature range.
Slotte	$\eta = \frac{a}{(b+T)^c}$	Reasonable; useful in numerical analysis.
Walther	$(\upsilon + a) = bd^{\left(\frac{1}{T}\right)^c}$	Forms the basis of the ASTM viscosity-temperature chart.
Vogel	$\eta = ae^{\frac{b}{(T-c)}}$	Most accurate; very useful in engineering calculations.

where:

a, b, c, d	are constants.
υ	is the kinematic viscosity [m^2 /s].
T	is the absolute temperature [K].

2.3. Viscosity pressure relationship

Lubricant viscosity increases with pressure. For most lubricants this effect is considerably larger than the effect of temperature or shear when the pressure is significantly above atmospheric. This is of particular importance in the lubrication of heavily loaded concentrated contacts which can be found, for example, in rolling contact bearings and gears. The pressures encountered in these contacts can be so high and the rate of pressure rise so rapid that the lubricant behaves like a solid rather than a liquid [10].

A number of attempts have been made to develop a formula describing the relationship between pressure and viscosity of lubricants. Some have been quite satisfactory, especially at low pressures, while others have been quite complex and not easily applicable in practice. The best known equation to calculate the viscosity of a lubricant at moderate pressures (close to atmospheric) is the Barus equation. The application of this equation to pressures above 0.5 GPa can, however, lead to serious errors. The equation becomes even more unreliable if the ambient temperature is high. The Barus equation is of the form:

$$\eta_p = \eta_0 e^{\alpha p} \tag{2.1}$$

where:

η_p is the viscosity at pressure p [Pas].

η_0 is the atmospheric viscosity [Pas].

α is the pressure-viscosity coefficient [m²/N], which can be obtained by plotting the natural logarithm of dynamic viscosity η versus pressure p. The slope of the graphs is α.

p is the pressure of concern [Pa].

For higher pressures Chu *et al.* (see for instance [10]) suggested that the following formula can be used:

$$\eta_p = \eta_0 (1 + C \times p)^n \tag{2.2}$$

where:

C, n are constants, n is approximately 16 for most cases and C can be obtained from a diagram.

The pressure-viscosity coefficient is a function of the molecular structure of the lubricant and its physical characteristics such as molecular interlocking, molecular packing and rigidity and viscosity-temperature characteristics.

There are many other formulae for viscosity-pressure relationships. These formulae allow for the calculations of viscosity changes with pressure under various conditions and to various degrees of accuracy.

An expression which is suitable for computing was initially proposed by Roelands and developed further by Houpert (see for instance [10]). To calculate lubricant viscosity at a specific pressure a particular form of the Barus equation was proposed:

$$\eta_R = \eta_0 e^{\alpha^* p} \tag{2.3}$$

where:

η_R is the viscosity at pressure p and temperature θ [Pas].

η_0 is the atmospheric viscosity [Pas].

α^* is the Roelands pressure-viscosity coefficient which is a function of both p and θ [m²/N].

p is the pressure of interest [Pa].

2.4. Viscosity-shear rate relationship

When the shear rate does not influence the kinematic viscosity of a fluid, then the fluid is called Newtonian. If, on the contrary, the kinematic viscosity depends on the shear rate, then the fluid is called non-Newtonian. Figure 2.1 shows the behaviour of some non-Newtonian fluids.

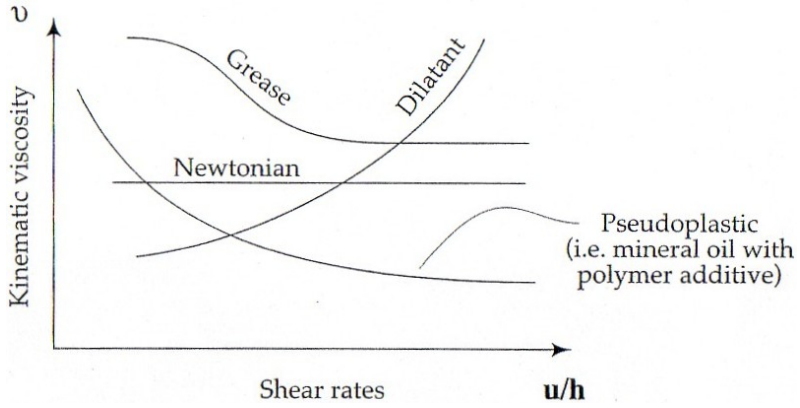


Figure 2.1 – Viscosity – shear rate characteristics for some non-Newtonian fluids [10].

Knowing the shear rate of a fluid at a specific shear rate is very important from the engineering point of view. For simplicity it is usually assumed that the fluids are Newtonian, i.e. their viscosity is proportional to shear rate. But in some cases, like the one of many lubricant oils and greases, it is necessary to find out a rheological model that describes the behaviour of these non-Newtonian lubricants as a function of the shear rate.

3. Mineral and synthetic oils

Mineral oils are the most commonly used lubricants mainly because of their low price. Even with the rapid development of synthetic oils, solid lubricants and wear resistant polymers, their continued use in many industries seems certain. There are certain advantages and disadvantages of applying mineral oil to lubricate specific machinery, and these must be carefully considered when selecting a lubricant and designing a lubrication system.

Despite many positive features such as availability and relatively low cost, mineral oils also have several serious defects, such as oxidation and viscosity loss at high temperatures, combustion or explosion in the presence of strong oxidizing agents and solidification at low temperatures. These effects are prohibitive in some specialized applications such as gas turbine engines where a high temperature lubricant is required, but occasionally very low temperatures must be sustained. In other applications such as vacuum pumps and jet engines, low vapour pressure lubricant is needed; in food processing and the pharmaceutical industry low toxicity lubricant is required, etc. In recent years the strongest demand has been for high performance lubricants, especially for applications in the aviation industry with high performance gas turbine engines. This led to the development of *synthetic oils* and lubricants that can withstand high temperatures without decomposing and at the same time will provide a reduced fire hazard. The recent trend towards high operating temperatures of machinery has created a second and probably more durable period of interest in these lubricants [10].

4. Greases

The American Society for Testing Materials (ASTM) defined lubricating grease as: *"A solid or semi-solid lubricant consisting of a thickening agent in a liquid lubricant. Other ingredients imparting special properties may be included".*

After this earlier definition, other attempts were made to redefine it. Nevertheless, and simplifying the question, greases are basically two-phase lubricants composed of a thickener which is physically dispersed in a base oil. The thickener is usually a soap or polymer fibres, which may vary in size from 1 to 100 microns, or dispersions of small (~ micron) particles. Greases therefore have a matrix structure; the scale of which can vary from simple particle interaction to the formation of extended fibre networks [67]. Since the oil is constrained and unable to flow it provides semi-permanent lubrication. For this reason, greases are widely used in spite of certain limitations in performance [10].

The most widespread application of grease is low-maintenance, semi-permanent lubricants in rolling contact bearings and some gears. The grease may be packed into a bearing or gear set and left for a period of several months or longer before being replaced. Inaccessible wearing contacts, such as are found on caterpillar track assemblies or in agricultural machinery, are conveniently lubricated by this means. Low maintenance items are also suitable candidates for grease lubrication. The lubricating performance of greases is inferior to mineral oils except at low sliding speeds where some greases may be superior. Greases have to meet the same requirements as lubricating oils but with one extra condition, the grease must remain as a semi-solid mass in spite of high service temperatures. If the grease liquefies and flows away from the contact then the likelihood of lubrication failure rapidly increases. Furthermore, grease is unable to remove heat by convection as oil does, so unlike oil, it is not effective as a cooling agent. It also cannot be used at speeds as high as oil because frictional drag would cause overheating. The lifetime of a grease in service is often determined by the eventual loss of the semi-solid consistency to become either a liquid or a hard deposit [10].

4.1. Composition

Lubricating greases consist of three basic elements, closely connected by surface mutual relations:

- Base oil (mineral or synthetic);
- Thickener (metal soaps and clays);
- Additives and fillers.

Base oils usually play the most important role in determining the grease performance, but also the type and amount of thickener (typically 5-20%) and the additives (similar to those in lubricating oils are used) can be critical.

4.1.1. Base oils

Both mineral oils and synthetic oils can be used as base oils in the manufacture of greases. Mineral oils are the most used ones, constituting the base oil type of approximately 90% of greases manufactured. Due to their relatively low price and good properties, mineral base greases are used to lubricate common mechanical parts. On the other hand, synthetic base greases, with better properties but also higher price, have their most common applications in high performance aircrafts, missiles and in space technology. They are designed to be fire resistant and to operate in extremes of low and high temperatures.

The viscosity of the base oil influences grease's consistency although, it is more dependent on the amount and type of thickener used.

4.1.2. Thickener

Grease characteristics depend on the type of thickener used. For example, if the thickener can withstand heat, the grease will also be suitable for high temperature applications; if the thickener is water resistant the grease will also be water resistant, etc. Hence the grease type is usually classified by the type of thickener used in its manufacture. As there are two fundamental types of thickener that can be used in greases, the commercial greases are divided into two primary classes: soap and non-soap based. Soap type greases are the most commonly produced and the most commonly used are calcium, lithium, aluminium, sodium and others (mainly barium) [10].

The thickeners form a soft, fibrous matrix of interlocking particles. The interlocking structure forms tiny pockets of about 10^{-6} m in which the oil is trapped. A representation of greases microstructure showing the fibrous structure of soap and macromolecules can be seen in Figure 2.2.

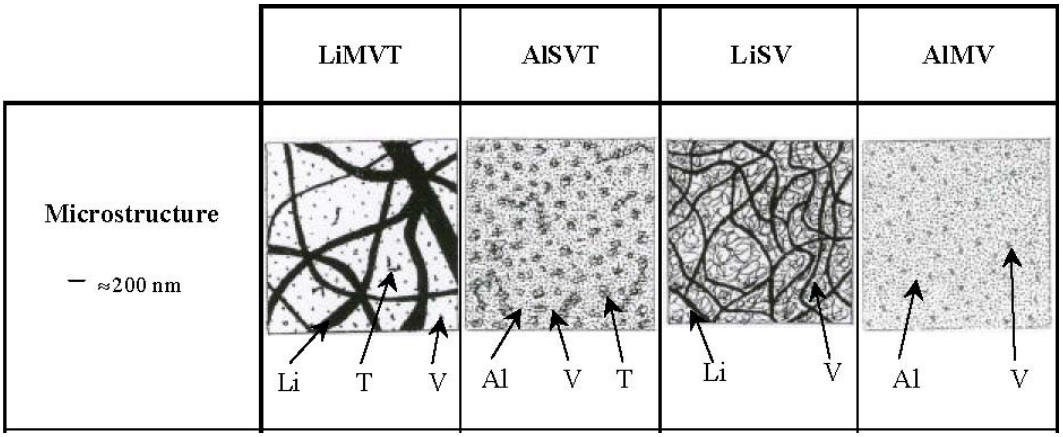


Figure 2.2 – Schematic of grease microstructure (soap and macromolecules) where M (mineral) and S (ester) represents the base oil type and V (viscosity improver) and T (adherence improver) represents the additives [69].

4.1.3. Additives and fillers

The additives used in grease formulations are similar to those used in lubricating oils. Some of them modify the soap, others improve the oil characteristics. The most common additives include anti-oxidants, rust and corrosion inhibitors, anti-wear (AW) and extreme pressure (EP) additives.

Fillers are sometimes used as fine solids in grease formulations to improve grease performance. Typical fillers are graphite, molybdenum disulphide, metal oxides and flakes, carbon black, talc and others.

4.2. Lubrication mechanism of greases

The lubrication mechanism of greases is still not fully understood. The question of how do greases lubricate and what is their lubrication mechanism still waits for an answer.

For example, in oil lubrication, the mechanism can be either hydrodynamic, elastohydrodynamic or boundary, depending on the operating conditions. The lubrication mechanism of greases, however, will be different since they have a different structure from oil. The structure of grease is gel-like or semi-solid. Baker (1958), Scarlett (1967) and Cann (1996) quoted three mechanisms for the formation of a lubricating film and for the replenishment of the lubricant inlet of the contact zone in a grease-lubricated rolling bearing [69]:

- The grease acts as a reservoir to provide a controlled supply of oil which bleeds in to the contact zone (Baker, 1958);
- The separation of the rolling elements and the raceways is due to a thin layer of highly worked grease deposited in the rolling track during the first few hours of running (Scarlett, 1967);
- The separating film consists of both thickener and base oil components (Cann, 1996).

This last mechanism described by Cann [70] showed that the bleeding of oil from the grease is not the principal mechanism of lubrication as it was thought earlier. It appears that the thickener as well as the oil takes part in the lubrication process, and that grease as a whole is an effective lubricant.

Cann and Lubrecht [71] have divided the resupply of the grease into the rolling contact into two mechanisms:

- The continuous replenishment mechanism: consists of the replenishment of the bulk grease in the rolling track and of the local reflow of base oil from grease reservoirs around the track.
- The intermittent replenishment mechanism: might consist of several mechanisms, which make the grease 'flow' in the bearing, and the vibration of the bearing is one of the most important factors for moving grease in the bearing.

The mechanism that dominates depends on the type of grease, the load and the design of the bearing. The efficiency of the replenishment mainly depends on the operational conditions of the bearing. It has been observed that, in angular-contact and thrust ball bearings, the spin motion of the balls is responsible for transporting the bulk grease into the rolling track. A strongly different situation

occurs, for example, in the cylindrical roller bearing in which the bulk flow of grease is minimal [69].

4.2.1. Film thickness in grease lubrication

Experimental studies in a ball-on-plate device using optical interferometry to measure EHD film thickness and the grease flow close to the contact have shown that, initially, the EHD films are much thicker than for the base oil. This is due to fresh grease being entrained forming what is called *fully flooded lubrication*. As the test proceeds grease is pushed aside by the overrolling of the ball and does not readily flow back to replenish the track. This can result in progressive bulk starvation of the inlet, and a significant drop in film thickness, leading to a situation of *starved lubrication*. This is unavoidable unless the supply is maintained by an external mechanism. If replenishment is externally maintained then film thickness depends on the effective grease viscosity in the inlet and degree of starvation [67].

The behaviour of the film thickness in grease lubrication above described is schematically illustrated in Figure 2.3.

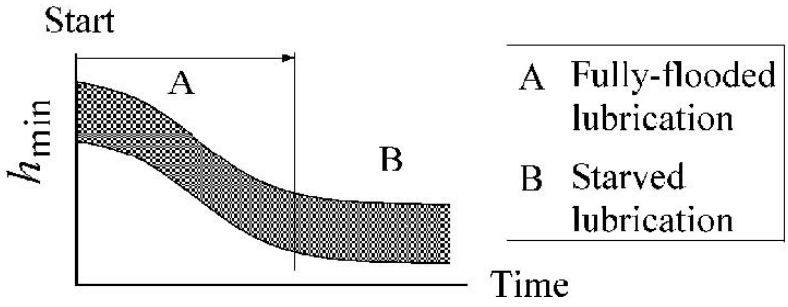


Figure 2.3 – Schematic illustration of the change in the film thickness in a grease-lubricated rolling contact with time [69]

Jonkisz and Krzeminski-Freda [71] and Aihara and Downson [72] have presented measurement results for grease film thickness in line contacts. Jonkisz and Krzeminski-Freda stated that in fully flooded conditions the grease film is about 1.5 times the oil-film thickness obtained with the base oil of the grease. Aihara and Downson concluded that the film thickness in grease lubrication at starved conditions is about 0.7 times the corresponding base oil film thickness. However, according to Åström *et al.* [73] the film thickness in starved grease lubrication is 0.35-0.45 times the base oil-film thickness. As a rough estimate, it can be assumed that the real value for the lubricating film thickness under starved conditions is between 35 and 70 per cent of the film thickness obtained with an oil of the same type as the base oil of the grease [69].

But what actually causes a drop in the film thickness drops changing a grease lubricated contact from fully-flooded to starved lubrication? Palacios *et al.* [74] have studied the formation of the film thickness by measurements in grease-lubricated point and line contacts. They measured that the film thickness in line contact decreased during the test time, which was a two hour long one. These authors explained that the drop in the film thickness in the film thickness was due to the rise in running temperature from 20 to 35 °C during the measurement. An additional explanation may be the bleeding of the base oil from the grease, or mechanical degradation of bulk grease, also causing increased oil bleeding [69].

4.2.2. Influence of grease composition on film thickness

Grease film thickness in fully-flooded and starved lubrication is highly influenced by the base oil viscosity, the % of soap concentration, the thickener type, the presence of additives, etc.

4.2.2.1. Under fully-flooded conditions

Base oil viscosity

Kendall and Williamson [75] concluded that an increase in the oil viscosity results in greater film thickness.

Kaneta *et al.* [76] concluded that the overall film thickness for the same thickener structure increases as the base oil viscosity increases, when neither starvation nor adhesion of the thickener takes place.

% of soap concentration

Kendall and Williamson [75] concluded that an increase in soap concentration results in greater film thickness.

Hurley and Cann [77] concluded that in high speed EHD lubrication regime grease film thickness increases with thickener concentration. This is shown in Figure 2.4 where the λ ratio is plotted for a range of lithium hydroxystearate greases at different temperatures. The ratio is 1 for the base oil (0% thickener) and with increasing thickener concentration the value increases. Typically, for each 1% of added thickener the ratio is increased by approximately 5% i.e. the film thickness increases by 5% relative to the base oil.

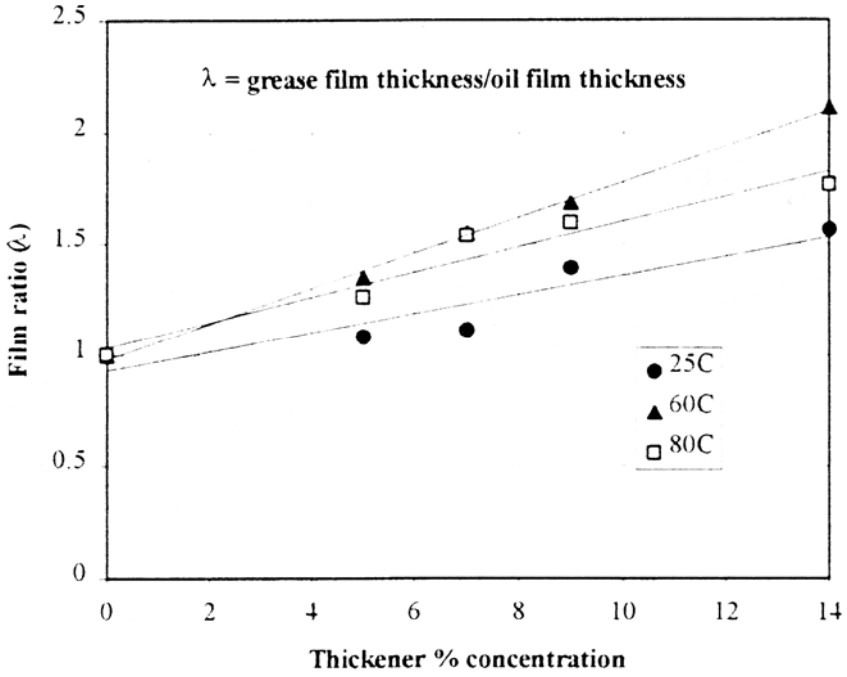


Figure 2.4 – Dependence of film ratio λ on thickener content for the lithium hydroxystearate greases [77].

4.2.2.2. Under starved conditions

Base oil viscosity

Cann and Lubrecht [78] concluded that the film thickness under starved conditions will be an interplay between the effect of viscosity on replenishment and loss and the effect on EHD lubrication film generation.

Generally, increasing the base oil viscosity and/or thickener content will increase the degree of starvation (defined as the starved film thickness divided by the fully flooded film thickness). Models of starved liquid lubricant systems propose that film replenishment in the track is inversely proportional to oil viscosity. Tests with spherical roller bearings showed that running time to failure decreased with increasing base oil viscosity [78].

% of soap concentration

Cann [79] has concluded that, in a starved lubrication situation, the degree of starvation increases with increasing base oil viscosity, grease thickener concentration and rolling speed. Correspondingly, the starvation decreases with increasing temperature. A higher thickener concentration reduces the bleeding of the base oil from the grease, hence the rolling contact becomes more starved.

4.2.3. Models of grease lubrication

It has proved difficult to incorporate the experimental observations into a generic model grease lubrication. Several film formation mechanisms have been offered for greases in both model and bearing contacts. The problem remains one of defining the supply mechanisms and the nature of the lubricant. Models have been suggested which encompass both the "continued flow" and "starved" behaviour observed in the experimental work. At present the only predictive models that have been developed are based upon fully flooded EHD theory [67].

The initial film thickness can be explained in terms of grease rheology [71]. The rheology of grease can be modelled by the Hershel-Bulkley equation:

$$\tau = \tau_p + \left(\eta_s \frac{du}{dh} \right)^n \tag{2.4}$$

where:

τ is the shear stress acting on the oil [Pa];

τ_p is the plastic flow stress [Pa];

η_s is the base oil dynamic viscosity [Pas];

$\frac{du}{dh}$ is the shear rate [s^{-1}];

n is a constant.

The value of n is close to 1. When n is exactly unity then the above equation reduces to the original Bingham equation which states that a fluid does not flow below a certain value of minimum shear stress. At high shear stresses, the fluid behaves as a Newtonian liquid. The Hershel-Bulkley equation usually gives a good agreement with experiment. When used in the theoretical analysis of EHD lubrication and compared with experimental results, good agreement between theoretical and experimental data has been obtained [10].

The models of grease lubrication are based in the Hershel-Bulkley equation and generally yield an expression relating grease film thickness to the rheological parameters. In its simplest form grease film thickness can be related to that of the base oil.

As it was referred above, the only predictive models developed up till now are based upon fully flooded EHD theory. This "full flow" approach does not, explicitly, offer a physical mechanism for grease lubrication however the assumptions inherent to the model have implications for the supply condition and

the nature of the film formed. Nevertheless, the “full flow” model approach is the only tool available at present for the prediction of grease film thickness in rolling contacts, although there are other physical models describing mechanisms of film formation where there is reduced, or negligible grease flow, into the contact.

Perhaps the most widely accepted view of grease behaviour in a bearing is that it acts as a reservoir releasing oil into the rolled track. Thus the ability of a grease to bleed oil is often taken as a guide to lubricating performance. This model implies that the lubricating film is composed of base oil alone and that its thickness is determined by the base oil properties and the efficiency of supply. There is little evidence, however, for the ‘oil bleed’ mechanism and another mechanism has been suggested in which a high viscosity layer is retained within the rolling track. The lubricant is formed of highly degraded grease which is deposited in the first few minutes of operation. Excess grease is pushed away from the rolling track and forms a close fitting seal to prevent lubricant loss.

Cann [67] stated that the dominant factor controlling grease film formation is the mechanism of lubricant supply to the contact. Furthermore, the nature of the material supplied and the efficiency of the process will determine the thickness and composition of the film formed.

4.3. Grease characteristics

There are several performance characteristics of greases which are determined by well established procedures. The most commonly used in the characterization of greases are consistency, drop point, evaporation loss, oxidation stability, apparent viscosity, stability in storage and use, colour and odour (see for instance Stachowiak and Batchelor [10]).

4.3.1. Consistency of greases

Consistency or solidity is a measure of the hardness or shear strength of the grease. It is defined in terms of grease penetration depth by a standard cone under prescribed conditions of time and temperature (ASTM D-217, ASTM D-1403). The consistency forms the basis for grease classification and its range is between 475 for a very soft grease and 85 for a very hard grease.

Although consistency is poorly defined it is a very important grease characteristic. The hardness of the grease must be sufficient so that it will remain as a solid lump adjacent to the sliding or rolling contact. This lump may be subjected to loads from centrifugal accelerations in rolling bearings and may also be subjected to frictional heat. However, if the grease is too hard 'channelling' may occur where the rolling or sliding element cuts a path through the grease and causes lubricant starvation. Excessively hard greases are also very difficult to pump and may cause blockage of the supply ducts to the bearings. Consistency of the grease also refers to the degree of aggregation of soap fibres. If the soap fibres are present as a tangled mass then the grease is said to be 'rough' and when the grease fibres have joined together to form larger fibres, the grease is said to be 'smooth'. Roughness or smoothness has a strong influence on the stable operation of rolling bearings. If the grease is too smooth, then will stable lumps of grease will never form in a rolling bearing, and high operating temperatures and short grease life will result. The trade term for this problem is that the grease has failed to 'clear'. For some unknown reason a very rough grease will be expelled from the bearing and the bearing will rapidly wear out. A grease that is neither too rough nor too smooth usually gives the lowest operating temperatures and least wear.

4.3.2. Mechanical stability

The consistency of a grease can change due to mechanical shearing. Even if at the beginning of the service grease possesses the optimum consistency for a particular application, mechanical working will damage the soap fibres and degrade the

grease. Greases differ significantly in the level of damage they will incur due to mechanical working. For example, greases working in gear boxes, bearings, or being pumped through pipes are subjected to shear. The changes in grease consistency depend on the stability of the grease structure. In some cases greases may become very soft or even flow, but in most cases there is only slight softening or hardening of the grease. Consistency of the grease is often specified for worked and pre-worked conditions. The grease is worked in the test apparatus which consists of a container fitted with a perforated metal plate plunger which is actuated by a motor driven linkage.

The consistency of greases made from several thickening agents has been measured after varying periods of mechanical working. It was found that all greases were softened by mechanical working to some extent, but when calcium tallow soap was the thickening agent, little damage resulted. Lithium hydroxystearate and sodium tallow stearate suffered significant damage initially, but thereafter their consistency reached a stable value. Lithium stearate and aluminium stearate, however, showed a continuous progression in damage.

It was also found that if the grease in a rolling bearing fails to clear then the continued mechanical working of the grease makes the situation even worse. The high operating speeds of rolling bearings accelerate the mechanical degradation of grease and it is advisable to operate the bearing at slightly less than the maximum rated speed. A design level of 75% of maximum rated speed has been suggested.

4.3.3. Drop point

The *drop point* is the temperature at which a grease shows a change from a semi-solid to a liquid state under the prescribed conditions. The drop point is the maximum useful operating temperature of the grease and it can be determined in an apparatus in which the sample of grease is heated until a drop of liquid is formed and detaches from the grease (ASTM D-566, ASTM D-2265). Drop point is commonly used as a quality control parameter in grease manufacturing.

4.3.4. Oxidation and thermal stability and evaporation loss

The *oxidation stability* of a grease (ASTM D-942) is the ability of the lubricant to resist oxidation. Oxidized grease usually darkens and acidic products accumulate in the same manner as is a lubricating oil. Acidic compounds can cause softening of the grease, oil bleeding, and leakage resulting in secondary effects such as

carbonization and hardening. In general the effects of oxidation in greases are more harmful than in oils.

Concerning the *thermal stability*, greases cannot be heated above a certain temperature without starting to decompose. The temperature limits for greases are determined by a number of grease characteristics such as oxidation stability, drop point and stiffening at low temperature.

As in oils, the *evaporation loss* of weight in greases due to evaporation can be quite significant. Volatile compounds and products of thermal degradation contribute to the losses, resulting in thickening of the lubricant, higher shear resistance and higher temperatures (ASTM D-972, ASTM D-2595).

4.3.5. Grease viscosity characteristics

Greases exhibit a number of similar characteristics to lubricating oils, e.g. they shear thin with increased shear rates, the apparent viscosity of a grease changes with the duration of shearing, and grease consistency changes with temperature.

Apparent viscosity of a grease is the dynamic viscosity measured at the desired temperature and shear rate (ASTM D-1092, ASTM D-3232) and it is useful in predicting the grease performance at a specific temperature.

Shear thinning of greases is associated with the changes in the apparent viscosity of grease with increased shear rates. When shearing begins the grease's apparent viscosity is high but with increased rates of shearing it may drop to that of its base oil.

Shear duration thinning of greases is associated with the changes which occur in the apparent viscosity of grease with the duration of shearing. Depending on the type of grease a permanent softening or reverse effect of hardening can occur. In some applications this effect can be beneficial, in others it is detrimental. For example, the permanent softening of a small quantity of grease in rolling contact bearings will result in good lubrication, low friction and low contact temperatures. On the other hand, the softening of the main bulk of grease will result in its continuous circulation and high operating temperatures.

Grease consistency temperature relationship describes the changes in the grease consistency with temperature. As has already been mentioned above the viscosity of oil is very sensitive to temperature changes. Relatively small temperature variations may result in significant changes in viscosity. There are only relatively small changes in grease consistency with temperature until it reaches its drop

point. At this temperature the grease structure breaks down and the grease becomes liquid.

4.4. Classification of greases

The most widely known classification of greases is related to their consistency and was established by the National Lubricating Grease Institute (NLGI). It classifies the greases into nine grades, according to their penetration depth, from the softest to the hardest [80], as shown in Table 2.2.

Table 2.2 – NLGI grease classification (adapted from [80]).

NLGI grade	Worked (60 strokes) penetration range [$\times 10^{-1}$ mm] at 25°C
000	445 – 475
00	400 – 430
0	355 – 385
1	310 – 340
2	265 – 295
3	220 – 250
4	175 – 205
5	130 – 160
6	85 – 115

Depending on the application a specific grease grade is selected. For example, soft greases, with NGLI grade 000, 00, 0 and 1, are used in applications where low viscous friction is required, e.g. enclosed gears which are slow, small and have a tendency to leak oil. In open gears grease must effectively be retained on the gear surface and tacky or adhesive additives such as bitumen are used in its formulation to improve adhesion. Greases with NGLI grade 0, 1 or 2 are used depending on the operating conditions such as speed, load and size of the gear. In rolling contact bearings greases with NGLI grade 1, 2, 3 and 4 are usually used. The most commonly applied is NLGI grade 2. Harder greases are used in large bearings and in applications where there are problems associated with sealing and vibrations. They are also used for higher speed applications. In plain, slowly moving bearings (1 – 2 [m/s]) greases with NGLI grade 1 and 2 are used. In general practice the most commonly used grease is the Multipurpose Grease

which is a grease with NGLI grade 2 according, with aluminium or lithium soap thickeners.

The selection of a grease for a specific application mainly depends on the temperature at which the grease is expected to operate. For low temperature applications the important factor is the low-temperature limit of a specific grease, which is determined by the viscosity or pour point of the base oil. The maximum operating temperature for a grease is limited by the drop point and the oxidation and thermal stability of the base oil and the other grease components. It is interesting to note that at temperatures above the drop point a grease may still provide effective lubrication but it will no longer be a grease since it will have changed its phase and become a liquid.

4.5. Degradation of greases

Even though grease is prone to a greater number of degradation modes than oil, it is required to spend a greater period of time as a functioning lubricant. Grease remains packed within the rolling bearing, gear, etc., whereas oil is circulated from a sump. Grease failure often does not occur immediately but small changes in operating conditions, particularly temperature, may cause problems associated with grease degradation. For example, at 40ºC the lifetime of a lithium hydroxyl-stearate grease is approximately 20,000 hours, whereas at 140 ºC its lifetime is only 500 hours. Grease failure in these circumstances is caused by hardening of the grease and formation of deposits on bearing surfaces [10].

5. Lubricant additives

Lubricant additives are chemicals, nearly always organic or organometallic, that are added to oils in quantities of a few weight percent to improve the lubricating capacity and durability of the oil. Specific purposes of lubricant additives are:

- improving the wear and friction characteristics by provision for absorption and extreme pressure (EP) lubrication;
- improving the oxidation resistance;
- control of corrosion;
- control of contamination by reaction products, wear particles and other debris;
- reducing excessive decrease of lubricant viscosity at high temperatures;
- enhancing lubricant characteristics by reducing the pour point and inhibiting the generation of foam.

Carefully chosen additives are extremely effective in improving the performance of an oil or grease. The most common packages of additives used in these lubricants contain anti-wear and EP lubrication additives, oxidation inhibitors, corrosion inhibitors, detergents, dispersants, viscosity improvers, pour point depressants and foam inhibitors. Sometimes other additives like dyes and odour improvers are also added to the oils [10].

Conclusions

The fundamental physical properties of lubricants which determine its lubrication and performance characteristics and the two types of lubricants (oils and greases) that can be found in the widest range of mechanical applications have been presented in this chapter. Particular emphasis was given to greases, the lubricants investigated in the experimental work. It was found that the lubrication mechanism of greases is still not fully understood contrarily to the case of oil lubrication.

Given the fact that the main goal of this thesis is to investigate the influence of grease composition on rolling contact fatigue damage, search in the available literature was performed trying to found studies on this subject, but nothing was explicitly found. Only studies concerning the influence of grease composition on film thickness were found and although some relations were established a considerable amount of research needs to be carried out in order to clearly understand the dependency of film thickness on grease composition.

The subjects presented in this chapter will stand as a support to explain the experimental results obtained in Chapter IV using greases with different compositions.

In the next chapter, the experimental methodology followed in order to test the different lubricants is presented.

Chapter III

Experimental methodology

1. Introduction

In this chapter the experimental procedures for the rolling fatigue tests and the means implemented to carry them out are presented.

For this study, six fatigue tests, under pure rolling conditions, were carried out in a twin-disc machine. The aim was to evaluate the influence of three types of greases and their correspondent base oils on the behaviour of the discs surfaces to RCF damage. The operating conditions were chosen so that an EHD contact similar to the one that can be found in rolling bearings was simulated.

The discs were manufactured in AISI 52100 steel, one of the most used in the manufacture of bearings, and followed by a heat treatment to obtain the desirable surface hardness. In each pair of discs, there is one cylindrical ("cylindrical surface") and another with a transversal curvature ("spherical surface"), in order to produce an ellipsoidal contact surface (see Figure 3.4). This procedure has the advantage of avoiding contact stress raise caused by edge effect, a phenomenon that can happen if there is some misalignment between two cylindrical discs. On the other hand, wear may alter the original curvature of the spherical disc, gradually changing contact conditions along tests [47].

Prior to the tests, the surface roughness of all the discs was measured and three artificial dents, separated by 120º, were performed on the surface's centre of each spherical disc. This artificial dents technique was used to accelerate the fatigue damage process and also served as a mean to compare the RCF damage of the discs for each lubricant tested and to calculate the wear volume. Some tests were performed in order to determine the load to be used for printing the artificial dents.

Surface analysis was made throughout the tests using a optical video-microscope device. After a certain number of fatigue cycles the discs were dismantled from the machine and the surfaces were photographed.

Sub-surface analysis was made at the end of the experimental tests using scanning electron microscopy (SEM) after cross-sections of the dents were performed. Perfilometry was also used to analyze the configuration of spall craters.

2. Experimental apparatus

2.1. Working principle of the twin-disc machine

The twin-disc machine was designed to reproduce the bearings working conditions, more specifically, the contact between the rolling elements and the raceway. The parameters values of the contact conditions are reproduced at the real scale and in a steady regime. Therefore, several aspects concerning this kind of contacts can be study, such as the materials and lubricants influence on the fatigue life of the contacting surfaces.

As it can be seen in Figure 3.1, the contact is made between two discs with parallel axis.

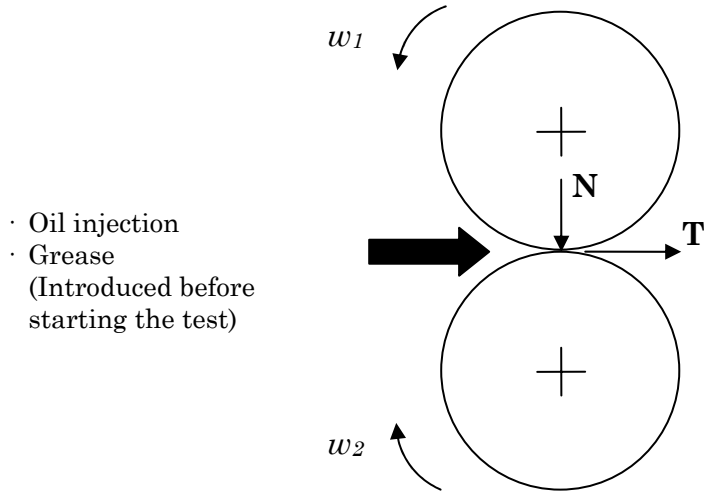


Figure 3.1 – Working principle of the twin-disc machine. *N* is the normal load and *T* is the sliding force.

The mechanical parameters ruling the contact are:

- the discs geometry,
- the discs micro-geometry,
- the linear speeds of the surfaces,
- the normal loading in the contact,
- the type of lubricant,
- the flow rate and the lubricant's temperature (when grease is being tested the temperature of the lubricant cannot be controlled).

In Table 3.1, the range of the mechanical parameters ruling the contact is presented.

Table 3.1 – Variation rates of the mechanical parameters ruling the contact.

rolling speed ω1, ω2 (rpm)	0 – 3000
rolling velocity U1, U2 (m/s)	0 – 11
slide roll ratios (U1-U2)/(U1+U2) (%)	0 – 24
normal load W imposed (N)	0 – 10000
maximum Hertz pressure P0 (GPa)	0 – 4
lubricant type	any
temperature of the lubricant (ºC)	any

2.2. Description of the twin-disc machine

The twin-disc machine used was manufactured in France, in the *"Contact and Solid Mechanics Laboratory"* of the INSA and was originally design for tests using oils. Later, in the scope of this work, a re-design of the machine was made so that the tests using greases became possible (the detailed description is presented in the Appendix A). The following figures (Figure 3.2 and Figure 3.3) show an overview of the twin-disc machine from the laboratory of the CETRIB – FEUP [81].

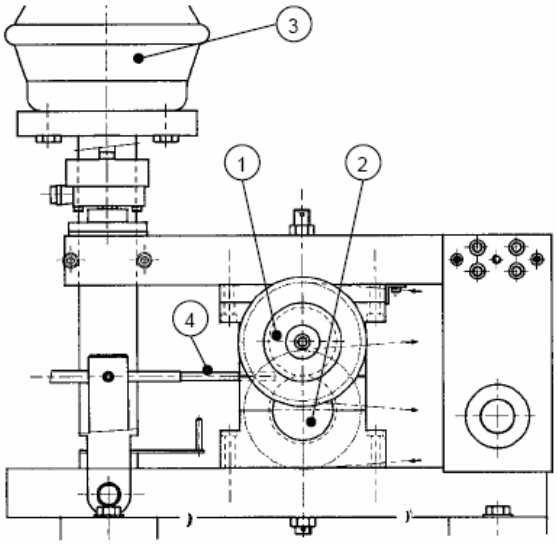


Figure 3.2 – Partial schematic view of the twin-disc machine.

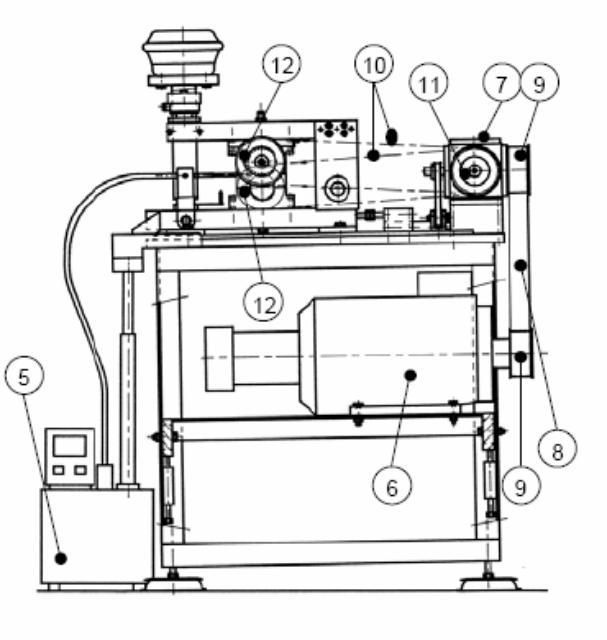


Figure 3.3 – Overall schematic view of the twin-disc machine.

Two discs (1,2) roll against each other, in conditions of pure rolling or rolling and sliding, transmitting a specified normal load, applied by a pneumatic cylinder (3). The discs are lubricated by oil injection (4) or by grease placed manually in the contact before running the experiments. The lubricating oil is kept at the specified temperature in the tank (5).

An electric AC motor (6), with frequency variation, drives an angle gearbox (7) by means of a toothed belt (8) and two toothed pulleys (9). Two other toothed belts (10) connect the toothed pulleys mounted on the disc shafts (11) to the toothed pulleys mounted on the angle gearbox output shafts (12).

The specified rolling and sliding speeds are obtained selecting the electric motor rotating speed and the adequate teeth numbers of pulleys (11) and (12).

Two proximity sensors measure permanently the distance that separates the surfaces of the two rolling discs. The information obtained from the sensors can be visualized in an oscilloscope localized in a central control system where a programmer automatizes the detection of superficial damages, as “spalling”. This control system also controls the temperature of the lubricant (when the tests are using oil) and the load in the pneumatic cylinder.

3. Characterization of the discs

3.1. Material

The material used in the manufacture of the discs was the AISI 52100 steel. As it was referred in the introduction, this is the most common material used for the manufacture of all type of bearings.

3.1.1. Chemical composition

In Table 3.2 the chemical composition of the AISI 52100 steel used in the manufacture of the discs is presented. As it can be seen this is a steel rich in chromium.

Table 3.2 – Chemical composition of the AISI 52100 steel.

Element	C	Si	Mn	Cr
Weight %	1,00	0,25	0,30	1,50

3.1.2. Heat treatment

The heat treatment that was performed in the AISI 52100 steel discs is described in Table 3.3. The structure of the AISI 52100 steel is martensitic and the hardness of the discs obtained after the heat treatment was approximately 58 HRc for all the discs.

Table 3.3 – Heat treatment of the AISI 52100 steel.

Operations		**Environment**	**Temperatures (ºC)**	**Time (min.)**
Tempering	1º Pre-heating	Air	≈ 200	15
	2º Pre-heating	Oven	600/650	30
	Heating	Salt bath	830	5
	Quenching	Oil	60	10
Annealing		Salt bath	240	30

3.2. Macro-geometry of the discs

Concerning the disc's macro-geometry, only the *crown radius* can be defined according to the preferences of the experimental work. The dimensions of the discs are fixed to a diameter of 70 mm and a thickness L of 7 mm as it is shown in Figure 3.4a. Therefore, the disc's radiuses are: $R_{x1} = R_{x2} = 35$ mm.

In each par of discs there is what can be named of a "spherical" and a "cylindrical" disc. The lower disc of the par is perfectly cylindrical because its crow radius is infinite ($R_{y2} = \infty$), while the upper disc is called spherical because its crown radius has a given value (in this study $R_{y1} = 35$ mm).

When the discs are loaded against each other, an elliptical contact area is generated, as it can be seen in Figure 3.4b. The dimensions of the ellipse are defined by the crow radius and the contact load, according to the Hertzian contact theory. For the maximum contact pressure chosen for this work $P_0 = 2$ GPa, the ellipse dimensions are: $a = 0.52$ mm and $b = 0.83$ mm.

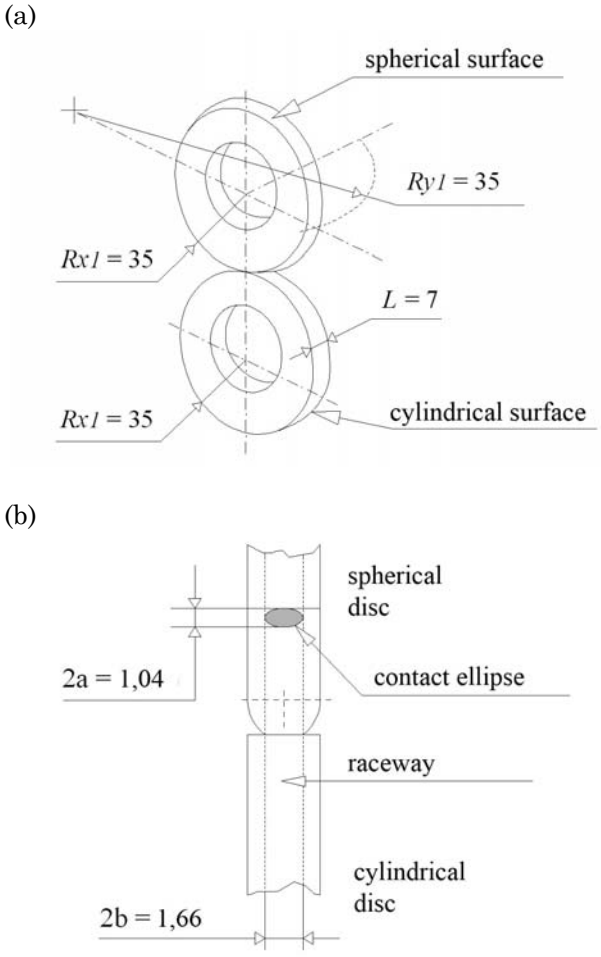


Figure 3.4 – Macro-geometry of the discs.

This geometry has some advantages: 1) it increases the contact pressure for a given load applied by pneumatic cylinder, 2) it eliminates the edge effect in the contact and 3) it eliminates the influence of a misalignment of the discs during the tests.

But, if the tests are long time tests, and/or involving high sliding rates and low materials ductility, then wear will occur in both surfaces. In the spherical surface, wear will increase the contact surface area and diminish the maximum contact pressure. Under these conditions, throughout the test, the contact geometry will tend to the one of a linear contact. This fact must be considered when evaluating the stresses field in the contacting bodies [47].

3.3. Micro-geometry of the discs

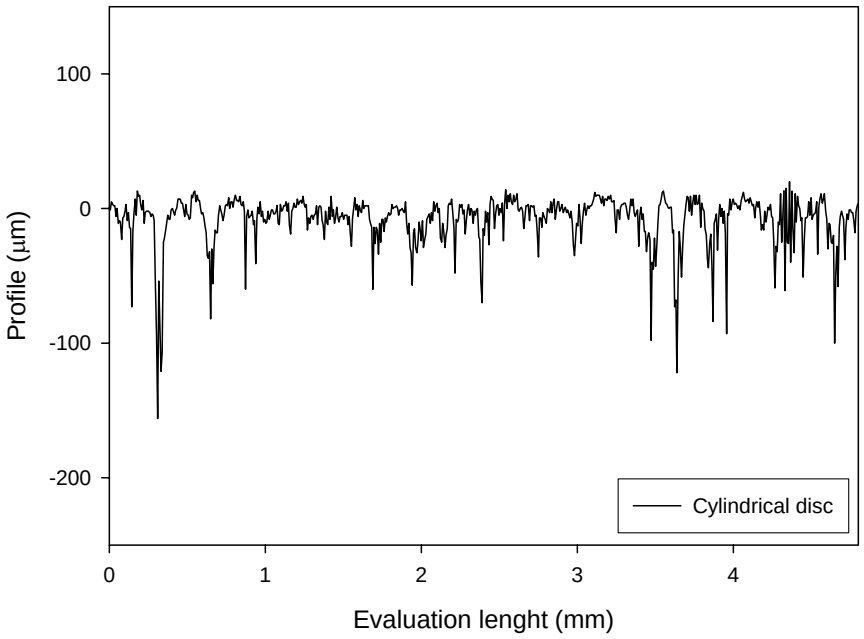
Table 3.4 shows the different roughness parameters measured in the surface of the discs using a *Hommelwerke Tester T 5000* roughness measurement device equipped with a mechanical sensor from CETRIB laboratory. An evaluation length of 4.80 mm and a cut-off of 0.8 mm were selected. The values of each parameter shown in the table correspond to an average of three different measurements that were performed for each disc's surface.

Table 3.4 – Roughness parameters measured.

Disc's reference		Roughness Parameters								
Pair	Type	Ra	RMS	R_{max}	$R_{z\text{-}d}$	R_k	R_{pk}	R_{vk}	M_{r1}	M_{r2}
1	C	0.155	0.210	1.820	1.445	0.390	0.085	0.395	6.050	81.40
	S	1.460	1.597	5.917	2.917	0.460	0.200	0.900	10.80	82.77
2	C	0.110	0.167	1.820	1.290	0.243	0.080	0.357	5.867	79.43
	S	1.430	1.603	4.683	3.083	1.107	0.200	1.050	5.467	83.07
3	C	0.090	0.140	1.503	1.150	0.213	0.053	0.333	5.400	80.93
	S	1.390	1.460	3.270	2.467	0.790	0.363	1.035	12.97	86.30
4	C	0.330	0.163	1.740	3.530	0.203	0.063	0.353	5.667	77.70
	S	1.397	1.497	2.843	2.230	0.467	0.093	1.015	5.883	85.95
5	C	0.150	0.067	1.933	1.067	0.300	0.070	0.400	1.233	83.53
	S	1.367	1.533	4.033	2.800	0.633	0.200	1.000	2.867	80.70
6	C	0.067	0.133	2.400	1.667	0.367	0.075	0.533	1.867	80.30
	S	1.267	1.400	3.100	2.400	0.633	0.200	1.067	5.633	81.47

In Table 3.4, for the disc's reference type, "C" means "Cylindrical" and "S" means "Spherical". As it can be seen both in Table 3.4 and in Figure 3.5a and 3.5b, the values of the different roughness parameters are considerably higher for the spherical discs because it is more difficult to obtain a better surface finishing in this case. This relatively higher roughness affects negatively the contact conditions and might enhance a higher wear of the contacting surfaces.

a)



(b)

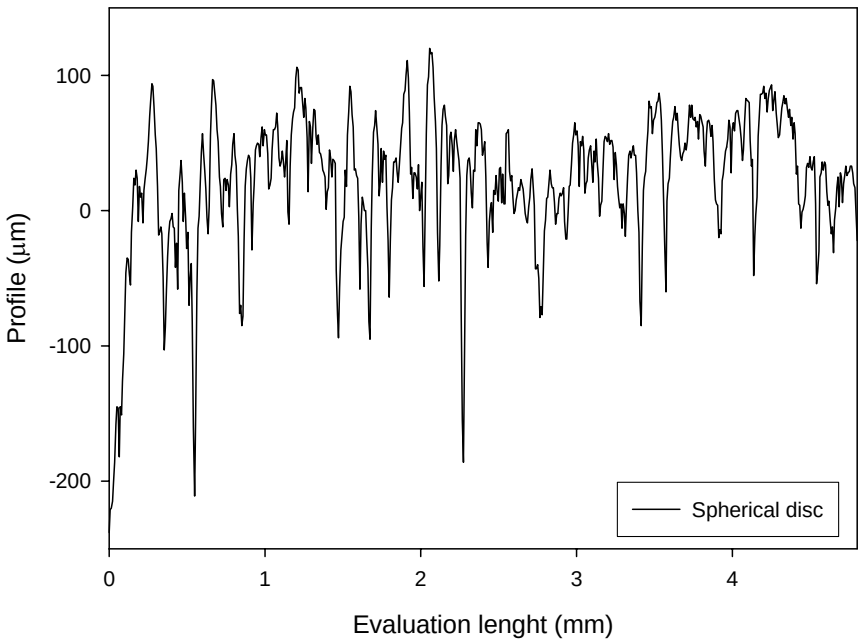


Figure 3.5 – Roughness profiles of a "cylindrical" (a) and a "spherical" (b) disc.

4. Artificial dents

Several authors (see Chapter I, Section 5) tested different types of artificial dents as a technique to concentrate local contact stress. After considering aspects as imposed area, stress level concentration and reproducibility, one particular form of surface dent was chosen: the hardness measurement Rockwell C. In fact, the dimensions of a Rockwell C dent can be controlled by the indent load, the stress level concentration is high due to a high plastic strain and these dents are easily reproducible. Furthermore, Rockwell C type dents are useful for modelling and calculation due to their symmetrical morphology (see Figures 3.6 and 3.7).

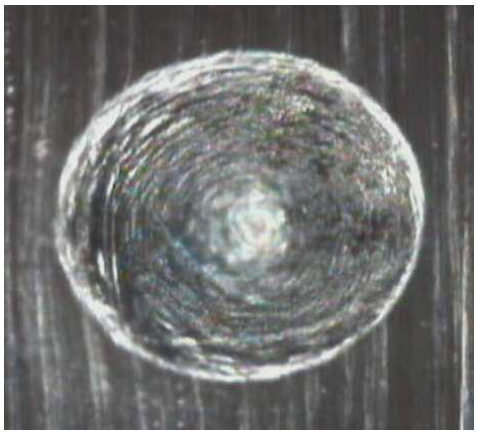


Figure 3.6 – Initial aspect of a Rockwell C dent

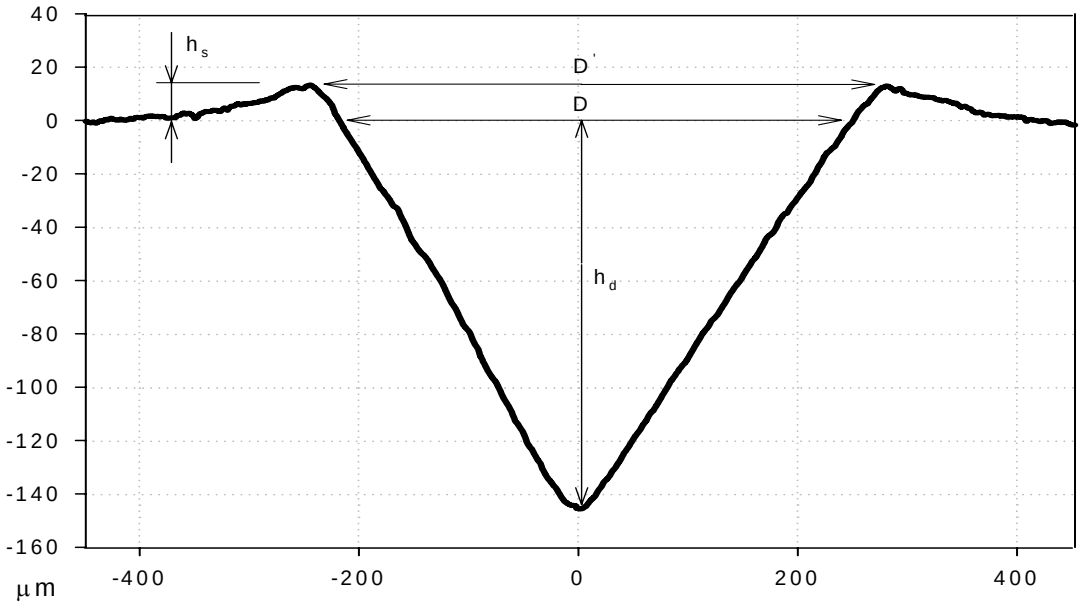


Figure 3.7 – Filtered profile of a Rockwell C dent.

To generate dents in the discs surface, a diamond Rockwell penetrator was manufactured and adapted to a tensile test machine from LET. The design of the penetrator is such that the location of the dent and the applied load should be controlled with great accuracy. The disc is fastened on a rotable and adjustable plate so that the penetrator could be pressed at the required location on the discs surface. The necessary indent load for getting the required dent can be adjusted by the load cell of the test machine.

To choose the adequate load, several tests were made in order to evaluate the dent size characteristics for different loads. The evaluated characteristics were the diameter D, the depth h_d, the shoulder height h_s and the shoulder sharpness S_s given by:

$$S_s = \frac{h_s}{\frac{D' - D}{2}} \tag{3.1}$$

The results, concerning the variation of these size characteristics for different indent loads, are presented, in Figure 3.8 to 3.11. As it can be seen, after 3000 N of applied load, the evolution tendency changes and although the depth and diameter grow, the shoulder height and consequently the shoulder sharpness decrease.

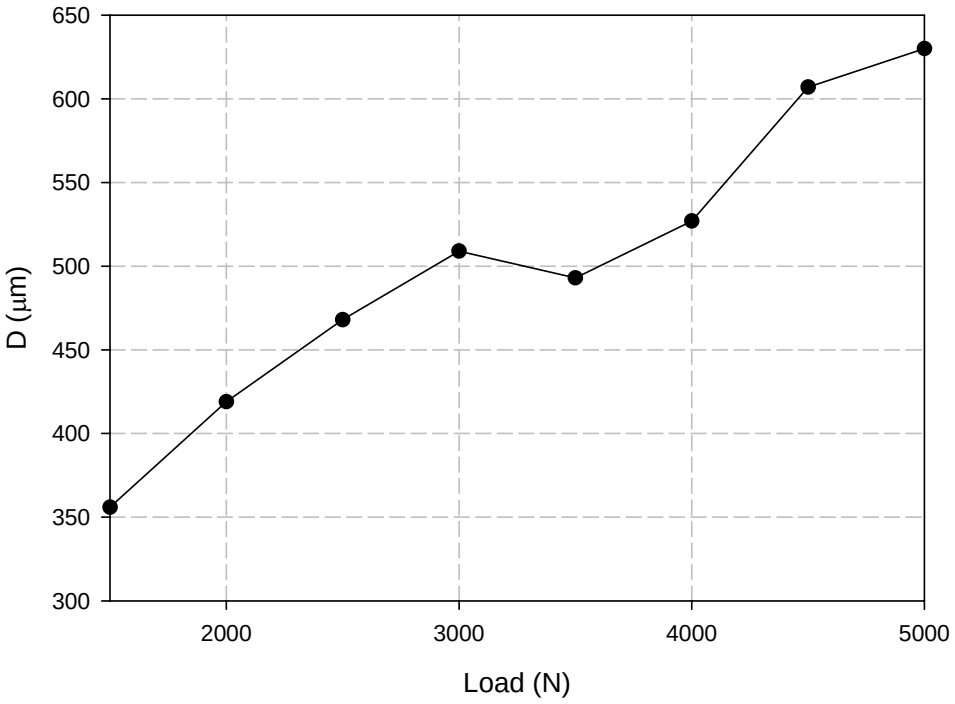


Figure 3.8 – Evolution of diameter D with the indent load.

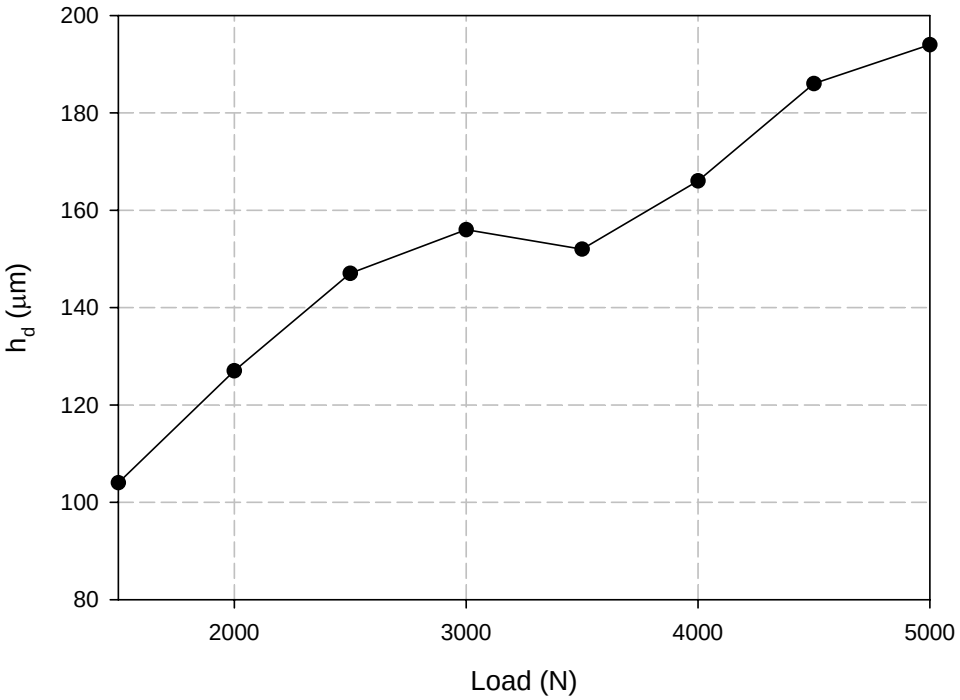


Figure 3.9 – Evolution of depth h_d with the indent load.

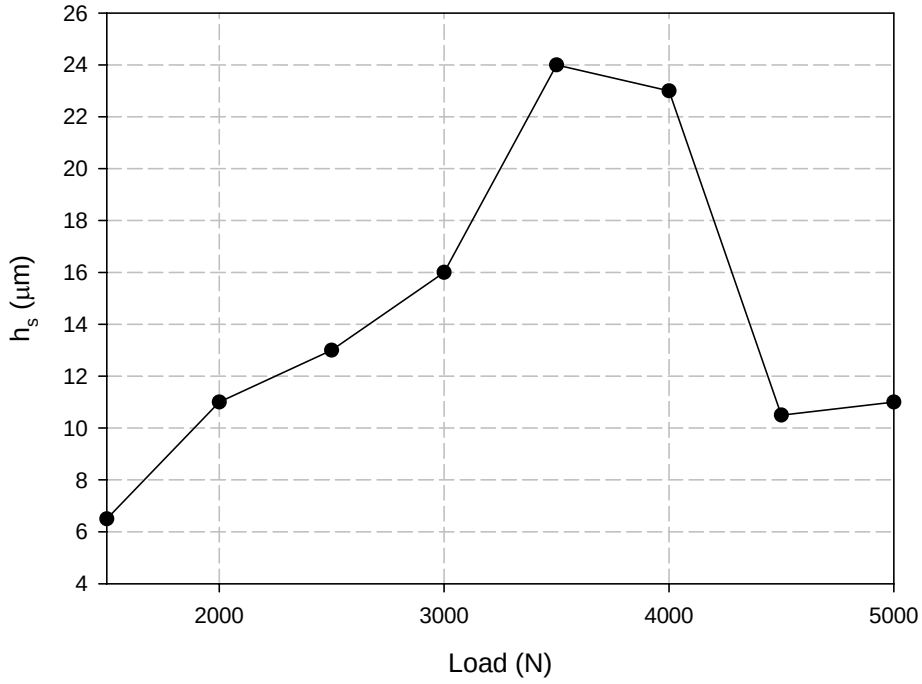


Figure 3.10 – Evolution of shoulder height h_s with the indent load.

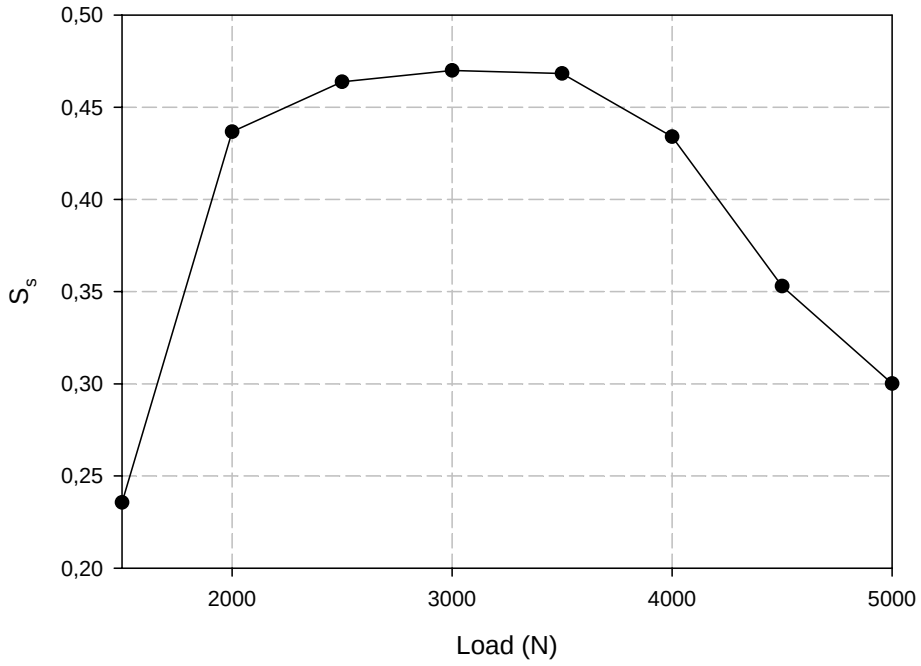


Figure 3.11 – Evolution of shoulder sharpness S_s with the indent load.

In Figure 3.12, a comparison between the different size characteristics is shown and in Figure 3.13 the different dent profiles for the different loads can be seen.

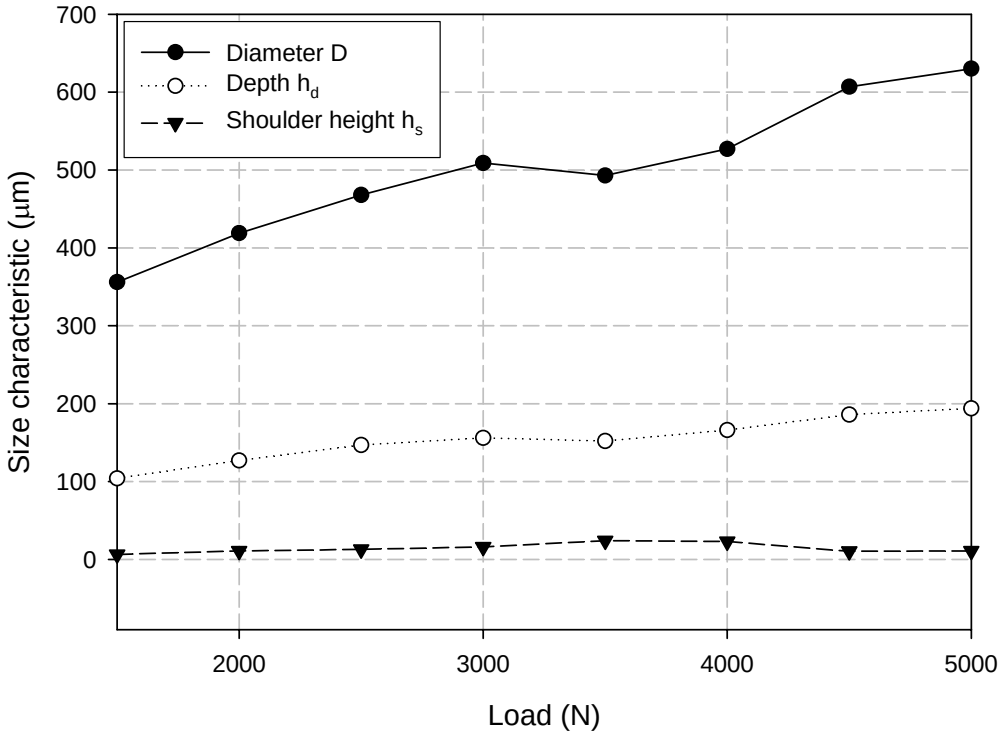


Figure 3.12 – Comparison of the size characteristics for different indent loads.

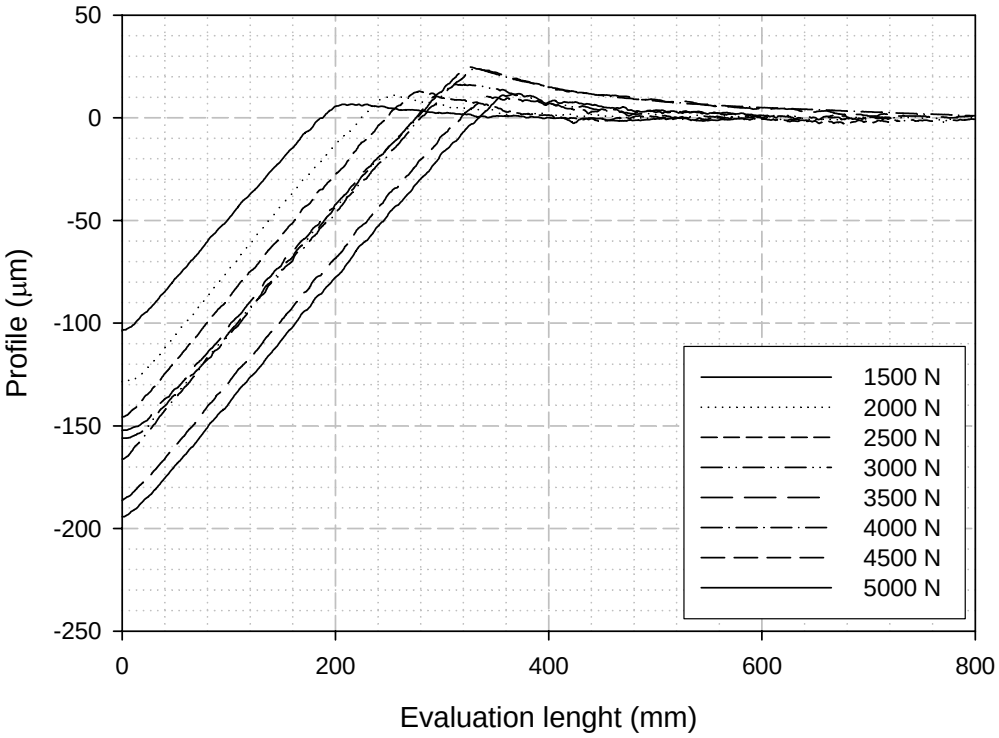


Figure 3.13 – Comparison of the dent profiles for different loads.

These tests at different loads helped in choosing the load that produces the most suitable dent for the fatigue tests. The best dent is the one that gathers the following characteristics:

- the smallest diameter (it must be as small as possible when compared with the contact ellipse);
- the lowest depth (a big depth might produce a exaggerated plastic deformation);
- a sufficiently high shoulder height and consequent shoulder sharpness (because the stress concentration is highly affected by the height of the dent's shoulder).

Analyzing the results for different loads, one can conclude that the preferable loads are the 2500 and 3000 N loads. The choice was the 2500 N load because, even having a slightly lower shoulder height and shoulder sharpness than the 3000 N load, it produces a dent with a smaller diameter (and also, but less significant, a smaller depth). This fact is very important because the dent area dimension must not approach in size the elliptical contact area.

After deciding about the indent load (2500 N), three dents separated 120º from each other were performed on the centre of the spherical surface discs of each pair. The reason for performing three dents is to ensure that at least one is exactly located in the centre of the contact.

5. Lubricants properties

Three greases and their respective base oils were tested. The main properties of these lubricants are shown in Tables 3.5 and 3.6.

Table 3.5 – Properties of the greases tested.

Properties	Greases		
	G1	G2	G3
Base oil	BO1	BO2	BO3
NLGI number	2	3	2
Thickener	lithium soap	lithium soap	lithium soap
Work penetration, 60 strokes (ASTM-D217)	265/295	220/250	265-295
Extended worked penetration, 100000 strokes (ASTM D-217) (max.)	+35	+35	+30
Dropping point, ºC (ASTM D-566) (min.)	185	185	185
Roll stability (ASTM D-1831)	unknown	unknown	7
Oxidation stability, psi drop/ 100 hours (ASTM D-942) (max.)	10	10	10
Oil separation, % (ASTM D-1742) (max.)	10	10	3
EP properties – Timken OK load test, lb (ASTM D-2782) (min.)	—	—	40
Cinematic viscosity of the base oils at 40 ºC, cSt (ASTM D-445)	100	100	219
Soap concentration, %	9	12	8.7

Table 3.6 – Properties of the base oils tested

Properties	Base Oils		
	BO1	BO2	BO3
Chemical form	Paraffinic	Naphthenic	Paraffinic
Cinematic viscosity at 40ºC (cSt)	100.000	124.495	530.873
Cinematic viscosity at 100ºC (cSt)	—	18.215	51.576
Specific gravity	0.8855	0.8968	0.8910

6. Experimental procedure

Six pairs of discs were tested in the twin-disc machine, one for each grease and correspondent base oil. For each test, exhaustive video-microscopic analyse was performed, following the evolution of the surface of the discs during the contact fatigue tests. Later, electronic microscopy and perfilometry were used to perform sub-surface post-test analysis, gathering information about different cracks that propagated from the discs surface.

6.1. Operating conditions

The different test conditions were the same for all the grease/base oil pairs (see Table 3.7). These conditions intend to represent those that can be found in mechanical applications such as gears and rolling bearings. In particular, a maximum Hertzian pressure of 2 GPa and a rolling speed of 3000 rpm represent extreme operating conditions that can be found in the referred applications.

Table 3.7 – Operating conditions

Material	AISI 52100 Steel
Geometry of the discs	$R_{x1} = R_{x2} = R_{y1} = 35$ mm $R_{y2} = \infty$ mm
Roughness of the discs surface	$R_a \approx 0.1$ (cylindrical discs) $R_a \approx 1.4$ (spherical discs)
Maximum Hertzian pressure	$P_0 = 2$ GPa
Maximum shear stress	$\tau_{max} = 0.65$ GPa
Depth of maximum shear stress	$Z_s = 315$ μm
Maximum orthogonal shear stress	$\tau_0 = 0.42$ GPa
Depth of maximum orthogonal shear stress	$Z_0 = 147$ μm
Rolling speed	$\omega_1 = \omega_2 = 3000$ rpm
Linear speed	$U_1 = U_2 \approx 11$ m/s
Sliding ratio	$SR = 0$ % (pure rolling)
Lubricant temperature (when oil was used)	$T_{oil} \approx 40$ ºC
Lubricant type	3 greases / base oils pairs

In Figure 3.14, the size of the indent produced by a 2500 N load is compared with the elliptical Hertzian contact area produced by the elastic deformation of the two discs. As it can be seen, the size of the dent is considerably smaller than the contact area.

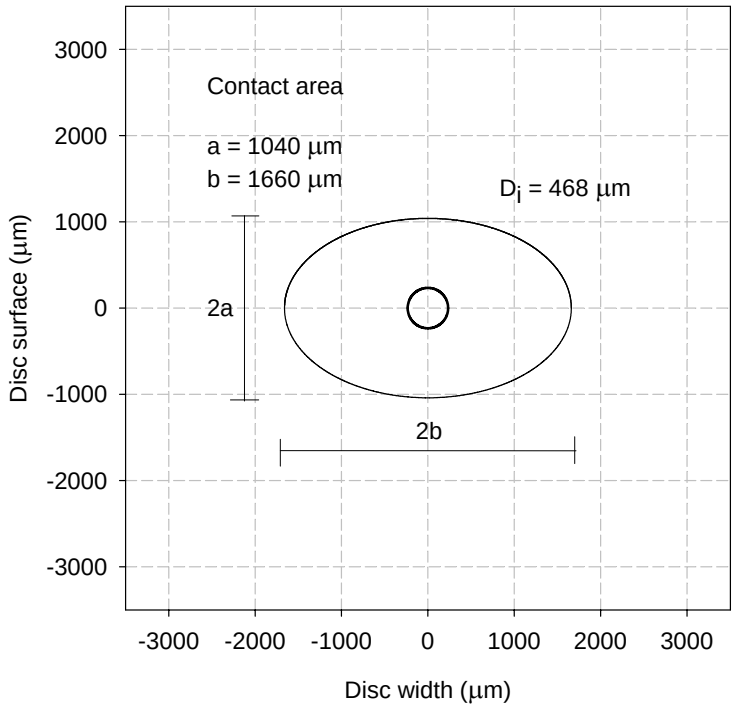


Figure 3.14 – Comparison between the contact area and dent dimension. D_i is the diameter of the dent; a is semiminor axis and b semimajor axis of the contact ellipse.

6.2. Tests and post-tests planning

Table 3.8 shows the test references and the variable parameters of each lubricant tested.

Table 3.8 – Test references and variable parameters for the tested lubricants

Test reference	Disc's pair	Variable parameters		
		Soap concentration	Viscosity of the base oil at 40ºC	Additives
G1	1	9 %	100.000 mm²/s	—
G2	2	12 %	124.495 mm²/s	—
G3	3	8.7 %	530.873 mm²/s	EP
BO1	4	—	100.000 mm²/s	—
BO2	5	—	124.495 mm²/s	—
BO3	6	—	530.873 mm²/s	—

All the tests had a total duration of 12 million cycles. To observe the increasing damaged area around the dents and the evolution of the surfaces condition, periodic stops were made. After every stop, the discs were dismantled from the machine and their surface was observed with an optical video-microscope using two objectives with a magnification of 100 times and 200 times.

In the first million of cycles, a higher number of stops was performed, in order to study the formation of the initial damaged area. After reaching 2 million cycles, stops were made every 2 million cycles.

Later, at the end of the tests, to get a more detailed observation some discs were cut into samples that were conveniently polished and treated in order to be observed using SEM. Both cutting and polishing processes were carefully performed, trying not to generate new cracks and preserving the metallic matrix condition. Cutting was done by a very slow electro-erosion process and the soft polishing removed at least a 0.5 mm thick, in order to ensure that cracks were to be observed resulted only from fatigue mechanisms and not from any post-test procedure.

6.3. Assessment of the severity of the contact

To assess the severity of the contact, the film thickness ratio λ was calculated for the different lubricants tested.

The film thickness ratio λ is the parameter characterizing the ratio of the minimum film thickness to the composite surface roughness (see Chapter I, Section 3.4) and is given by:

$$\lambda = \frac{\Phi_T h_0}{\left(\sigma_A^2 + \sigma_B^2\right)^{0.5}} \tag{3.2}$$

where:

h_0 is the minimum film thickness [m].

σ_A is the RMS surface roughness of body 'A' [m].

σ_B is the RMS surface roughness of body 'B' [m].

Φ_T is the thermal effects correction factor.

The film thickness for the base oil was calculated using the Hamrock and Dowson equation already presented in Chapter I, Section 3.3.2.

Greases film thickness calculation was made for fully-flooded and starved conditions using the following equations (see Chapter II, Section 4.5.1):

- fully-flooded conditions $\Rightarrow h_0 = 1.5 h_0^{baseoil}$ (3.3)
- starved conditions $\Rightarrow h_0 = 0.7 h_0^{baseoil}$ (3.4)

The results of the contacts severity assessment for the greases and correspondent base oils are presented in Table 3.9.

Table 3.9 – Contact severity for the greases and correspondent base oils.

Lubricant	Lubricating conditions	h_0 (μm)	σ (μm)	λ	Lubrication regime
G1	Fully-flooded	4,303	1,611	1,660	Full film
	Starved	2,008	1,611	0,775	Boundary film
G2	Fully-flooded	5,374	1,612	1,887	Full film
	Starved	2,508	1,612	0,881	Boundary film
G3	Fully-flooded	27,397	1,467	4,913	Full film
	Starved	12,785	1,467	2,293	Full film
BO1	Fully-flooded	2,869	1,506	1,184	Mixed film
BO2	Fully-flooded	3,583	1,534	1,321	Mixed film
BO3	Fully-flooded	18,264	1,406	3,416	Full film

Conclusions

In this chapter the experimental methodology was presented. The twin-disc machine, originally design to test oils had to be re-designed in order to test greases.

The type of artificial dents used was decided based on tests performed with different loads. It was found that a load of 2500 N produced the most efficient dent.

The lubricants used exhibit different percentages of soap concentration and base oils viscosity in order to test the influence of these properties on rolling contact fatigue.

Calculations to assess the severity of the contact revealed that under fully-flooded conditions, tests using greases always are under full film lubrication regime. It is expected to be so because the re-design of the twin-disc machine included a reservoir that allows replenishment of grease to the contact during the tests (see Appendix A).

In next chapter the results following this methodology are presented.

Chapter IV

Experimental results

1. Introduction

In this chapter, the experimental results obtained from the fatigue tests performed and the consequent analyses are presented.

The results are divided into two parts: surface and sub-surface observations. Surface observations were performed using an optical video microscope device while sub-surface observations were performed after making cross-sections of the discs through the dents using an electronic microscope.

Surface observations were used to characterize the damages on the dents and on the remaining surface of the discs and also to calculate the damaged area growth around the dents and the wear volume of the surfaces.

Sub-surface observations provided complementary information on fatigue damaged characterization around the dents.

The conclusions obtained from these results reveal new aspects concerning the influence of grease composition on rolling fatigue damage.

2. Surface observations

To analyse the surface damages the nomenclature proposed by Tallian (see Chapter I, Section 4) is adopted, where two main types of RCF are identified: 1) spalling, which is a failure by the formation of macroscopic craters in the contact surface and 2) surface distress, which is defined as an asperity-scale spalling fatigue.

The observed changes are reported throughout the fatigue cycles of each test. These observations were performed for the three dents printed in the spherical disc of each discs test. Different results for the dents were found and it was considered that the dent that better illustrates the results of each test is the one that suffered a more severe damaged.

All the pictures were taken with a magnification of 200 times except those in which a 100 times magnification is referred. No scalebar is presented on the pictures because this report only intends to show a qualitative analysis. For all the pictures concerning the observations on the spherical surface, the direction of the movement and load is the following (Figure 4.1).

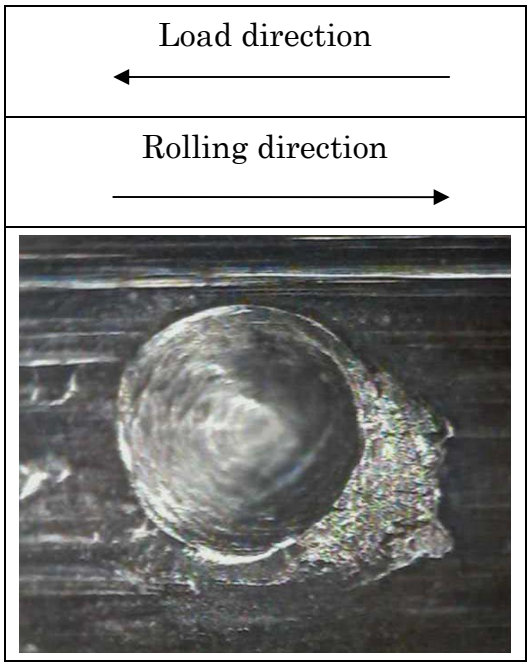


Figure 4.1 – Load and movement directions.

The following sections present an exhaustive report on the surface observations for all the lubricants tested.

2.1. Surface observations using greases

2.1.1. Using grease G1

Observations on the dents

- After 250,000 cycles: the dents shoulder was still visible.
- After 500,000 cycles: the dents shoulder was severely reduced. The first micro-cracks appeared around the dent forming a damaged area.

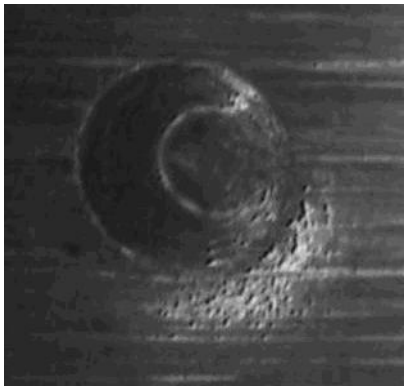


Figure 4.2 – Observations on a dent after 500,000 cycles using grease G1.

- After 2,000,000 cycles: the damaged area around the dents disappears due to wear produced.
- After 2,500,000 cycles: a damaged area reappears around the dents.
- After 3,000,000 cycles: a perfectly damaged area is visible next to the dents with a propagated crack that suggest the imminent appearance of a spall.

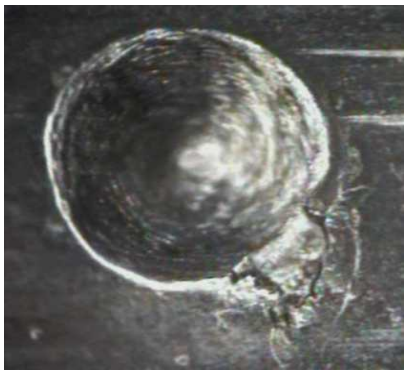


Figure 4.3 – Observations on a dent after 3,000,000 cycles using grease G1.

- Between 3,000,000 – 5,000,000 cycles: progressive growing of the damaged areas around the dents.
- After 5,500,000 cycles: the damaged areas grown considerably forming a clearly visible spall as it has been predicted from the observations after 3,000,000 cycles.

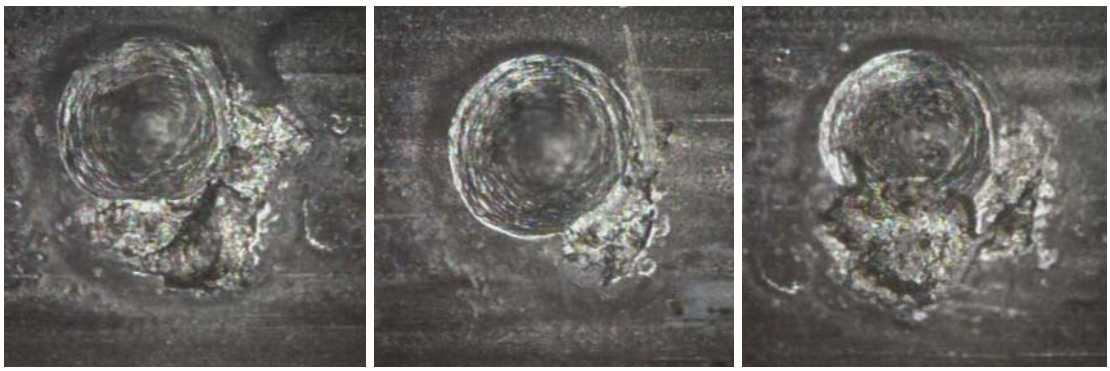


Figure 4.4 – Observations on the dents after 5,500,000 cycles using grease G1.

- Between 5,500,000 – 11,000,000 cycles: progressive growing of the damaged area around the dents.

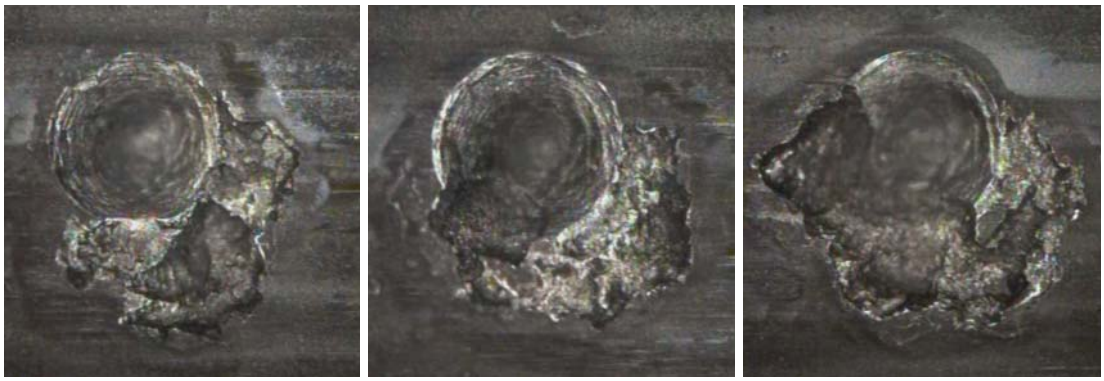


Figure 4.5 – Observations on the dents after 11,000,000 cycles using grease G1.

- After 12,000,000 cycles: the original dents shape is almost unrecognizable due to the considerable damage that occurred. A considerably large spall almost covers the original place of the dent. The area surrounding the dents is also considerably damaged.

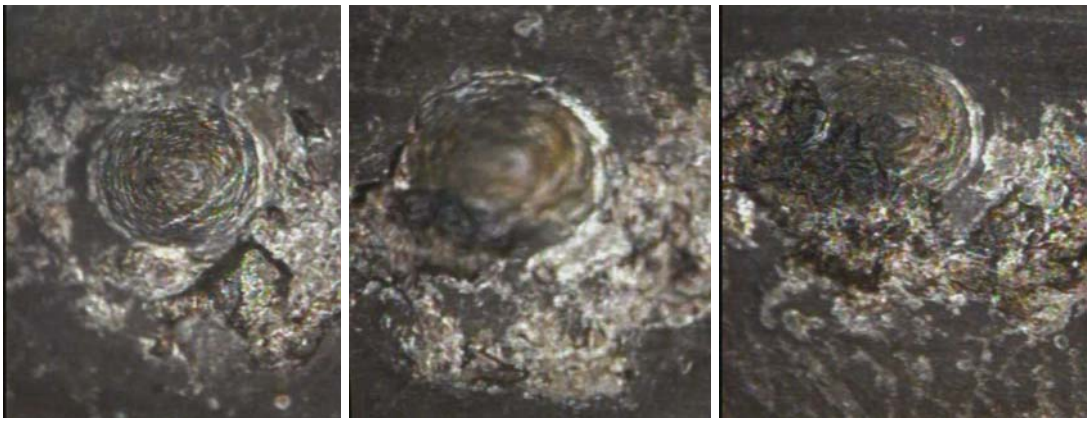


Figure 4.6 – Observations on the dents after 12,000,000 cycles using grease G1.

Observations on the remaining surface

- After 5,000,000 cycles it was observed that the surface became smoother due to gradual surface wear
- Natural dents: after 3,000,000 cycles natural dents were observed all over the surface of the discs and near the artificial dents. These natural dents are a consequence of the debris particles removed from the surface due to wear but also, and probably with a major influence, from the artificial dents surroundings. These metallic particles got into the contact interface and created a debris dent on the surfaces.

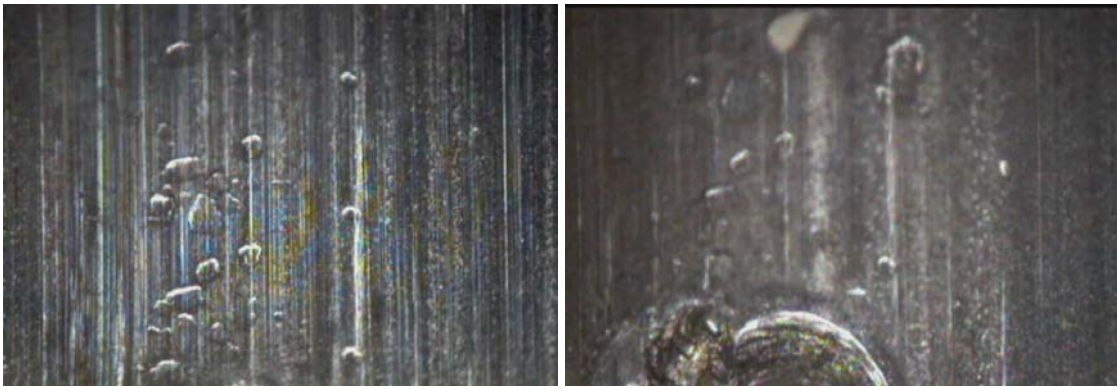


Figure 4.7 – Natural dents observed on the remaining surface after 3,000,000 cycles using grease G1.

During the experimental test it could be observed that the dimension of these natural dents growed.

After 5,500,000 cycles another type of natural dents was observed, characterized by its similar shape along the surface. Observation suggests that these dents are the result of particles that were grasped to the

cylindrical surface due to the Hertzian pressure in the contact. When contacting with the spherical surface, these particles printed successive dents. Also these dents dimension growed throughout the tests.

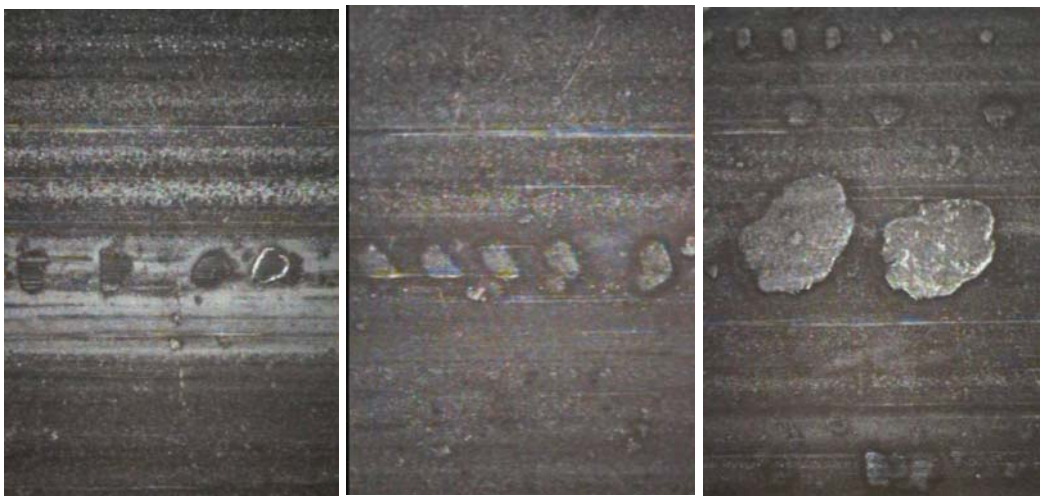
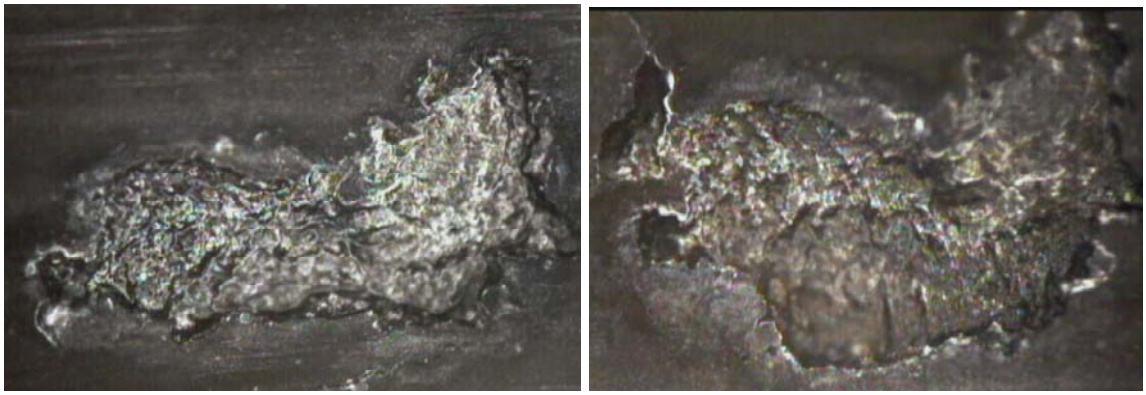


Figure 4.8 – Natural dents observed on the remaining surface after 3,000,000 cycles using grease G1.

Figure 4.9 – Natural dents observed on the remaining surface after 3,000,000 cycles using grease G1 (magnification of 100 times).

- Spall on the cylindrical disc: after 4,500,000 cycles the formation of a spall was observed on the surface of the cylindrical disc. Given the fact that this spall has considerable dimensions its appearance can be due to particle that got into the contact and dented the surface. Throughout the cycles the dimensions of this spall growed.



4,500,000 cycles 10,000,000 cycles

Figure 4.10 – Spall observed on the cylindrical disc using grease G1.

2.1.2. Using grease G2

Observations on the dents

- After 500,000 cycles: the dents shoulder was still visible although a small damaged area started appearing around the dents.

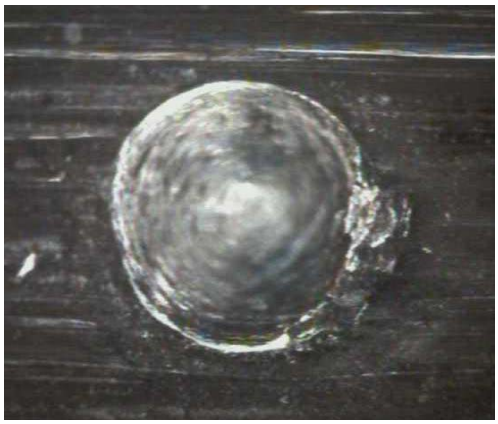


Figure 4.11 – Observations on a dent after 500,000 cycles using grease G2.

- After 1,000,000 cycles: the damaged area considerably increases and micro-cracks are perfectly visible.

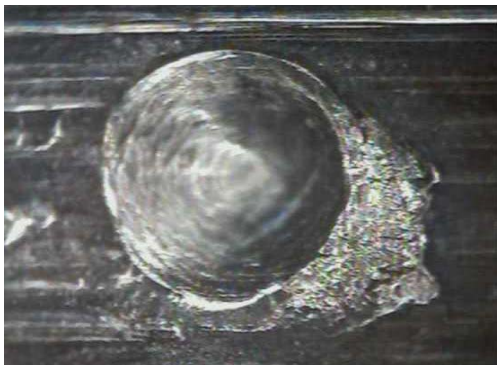


Figure 4.12 – Observations on a dent after 1,000,000 cycles using grease G2.

- After 2,000,000 cycles: the damaged area has increased considerably and some spalls can be distinguished.

Figure 4.13 – Observations on a dent after 2,000,000 cycles using grease G2.

- After 2,500,000 cycles: progressive damaged around the dents was observed. This is more visible for the dent that appears on the left picture, maybe due to the fact that this dent was right in the center of the discs, which is where the maximum Hertzian pressure occurs.

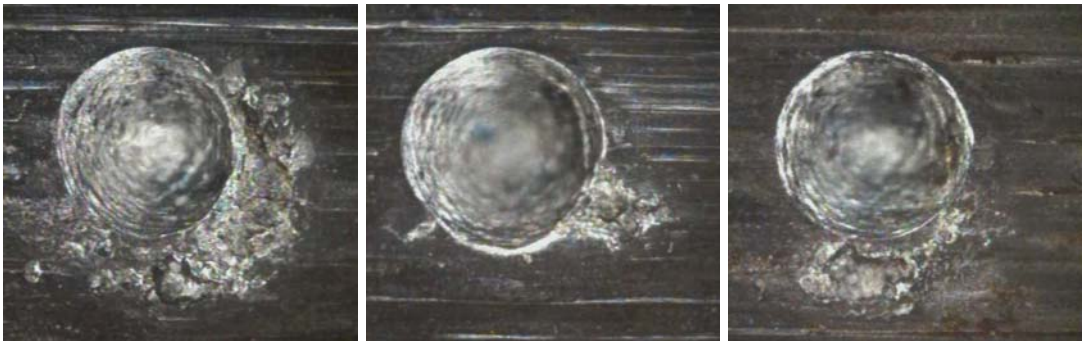


Figure 4.14 – Observations on the dents after 2,500,000 cycles using grease G2.

- Between 2,500,000 – 12,000,000 cycles: progressive growing of the damaged areas around the dents was observed.

  Surface distress was also observed after 3,500,000 cycles, with a severity that increased along the cycles.

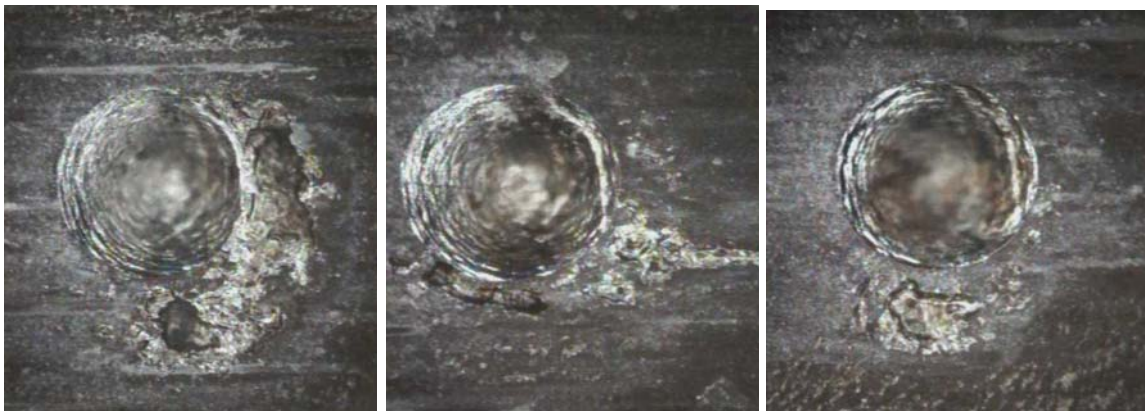


Figure 4.15 – Observations on the dents after 12,000,000 cycles using grease G2.

Observations on the remaining surface

- After 6,000,000 cycles it was observed that the surface became smoother due to gradual surface wear.
- Spall on the cylindrical disc: after 500,000 cycles, like it happened in the test using grease G1, the formations of a spall was observed on the surface of the cylindrical disc. The causes appear to be same, that is, the introduction of a debris particle in the contact. Nevertheless, this spall appeared at an early stage, probably due to a particle of bigger dimensions that was released from the one of the artificial dents. Throughout the cycles the dimensions of this spall also growed.

250,000 cycles 4,500,000 cycles

Figure 4.16 – Spall observed on the cylindrical disc using grease G2.

2.1.3. Using grease G3

Observations on the dents

- Between 500,000 – 2,000,000 cycles: the dents shoulder was still visible although a small damaged area started appearing around the dents.

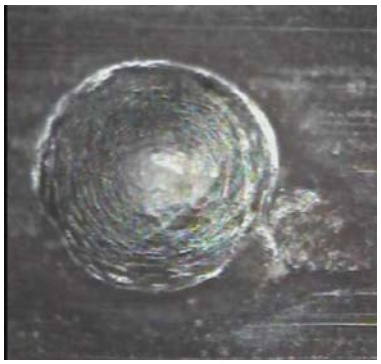


Figure 4.17 – Observations on a dent after 2,000,000 cycles using grease G3.

- Between 2,000,000 – 4,000,000 cycles: progressive growing of the damaged area around the dents was observed.

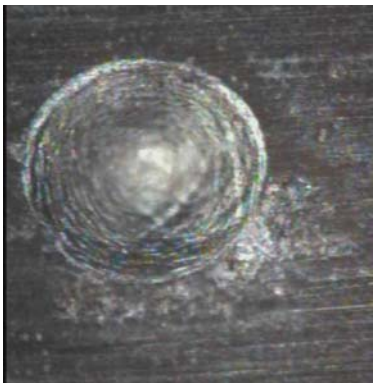


Figure 4.18 – Observations on a dent after 4,000,000 cycles using grease G3.

- After 5,000,000 cycles: a more significant damage around the indents was observed although no spalling occurred.

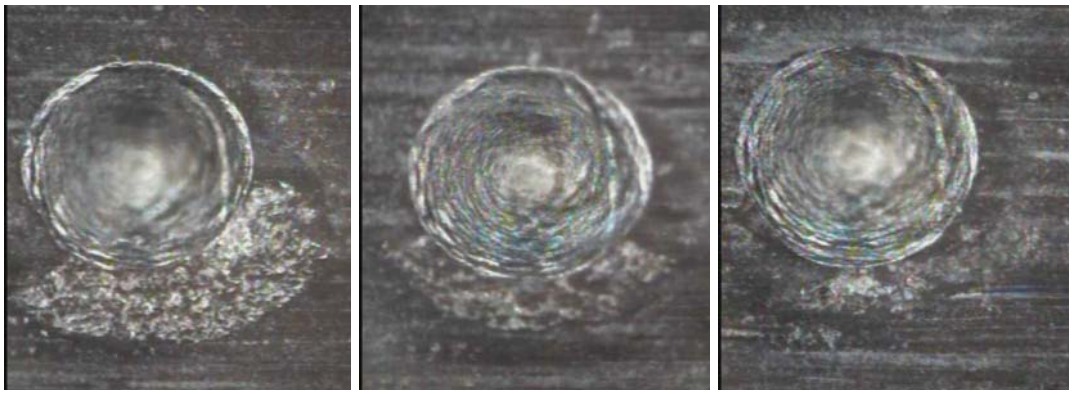


Figure 4.19 – Observations on the dents after 5,000,000 cycles using grease G3.

- Between 5,000,000 – 12,000,000 cycles: The damaged area remained more or less constant and no growing was observed. Contrarily, a reduction of the micro-cracks amplitude was observed due to gradual surface wear. At the end of the test spalling was not observed.

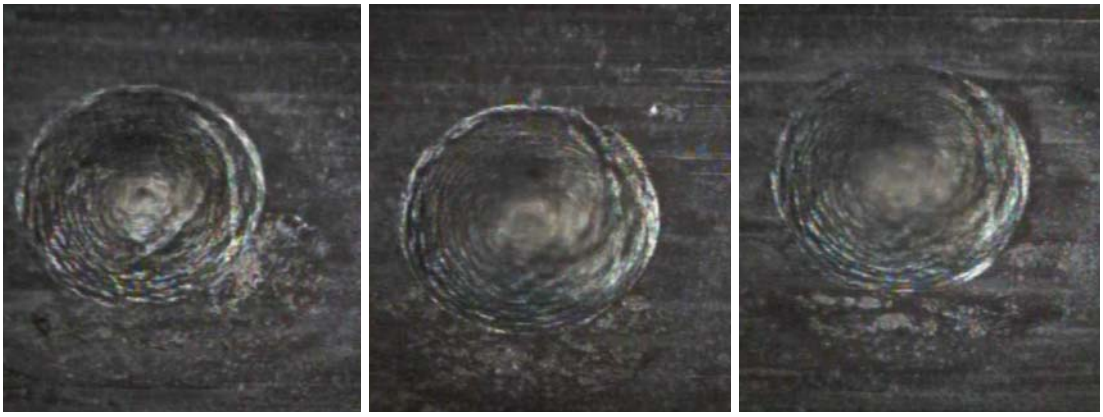


Figure 4.20 – Observations on the dents after 12,000,000 cycles using grease G3.

Observations on the remaining surface

- After 6,000,000 cycles it was observed that the surface became smoother due to the gradual surface wear.

2.2. Surface observations using base oils

2.2.1. Using base oil BO1

Observations on the dents

- After 1,000,000 cycles: the dents shoulder was still visible and a small damaged area appeared.

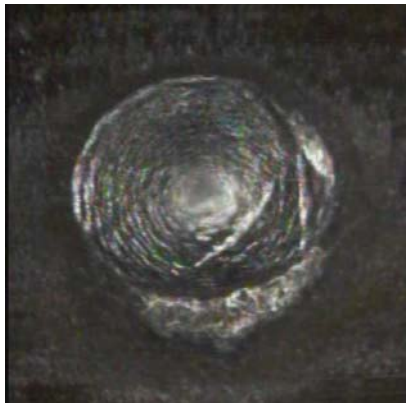


Figure 4.21 – Observations on a dent after 1,000,000 cycles using base oil BO1

- After 2,000,000 cycles: a generalized damaged area around the dents was observed.

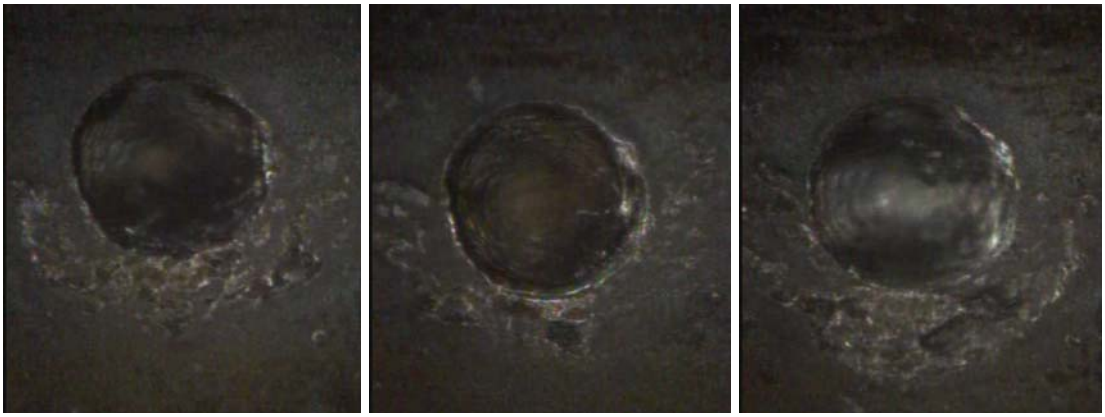


Figure 4.22 – Observations on the dents after 2,000,000 cycles using base oil BO1.

- After 3,000,000 cycles: the damaged area became generalized around the dents, revealing the appearance of surface distress.

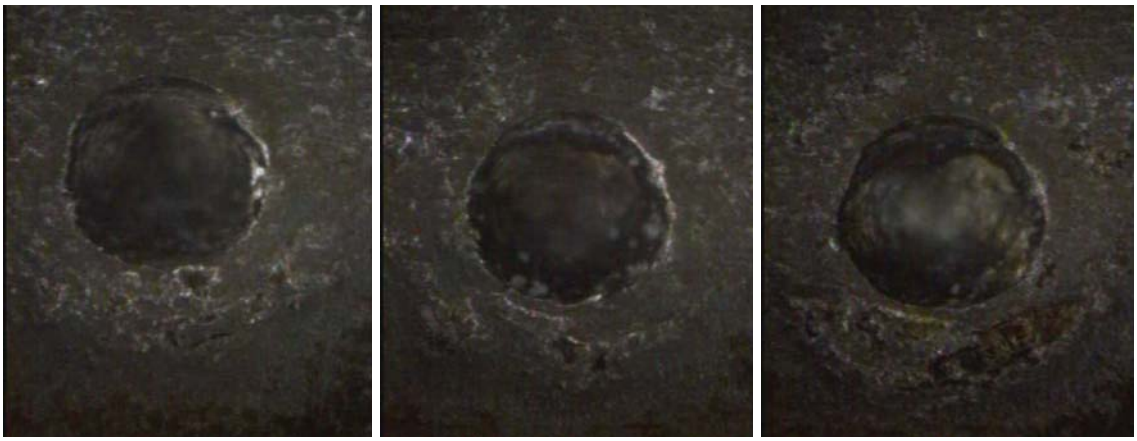


Figure 4.23 – Observations on the dents after 3,000,000 cycles using base oil BO1.

- Between 3,000,000 – 12,000,000: the generalized damaged area around the dents increased gradually and micro-spalls were observed next to the dents.

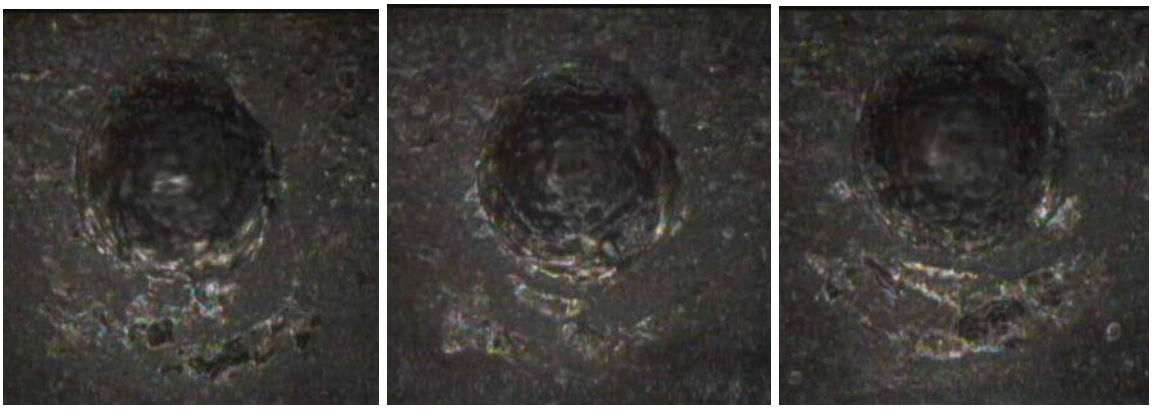


Figure 4.24 – Observations on the dents after 12,000,000 cycles using base oil BO1.

Observations on the remaining surface

- After 1,000,000 cycles it was observed that the surface roughness became smoother due to the gradual surface wear.
- After 10,000,000 cycles, generalized surface distress and some micro-spalls were observed on the remaining surface.

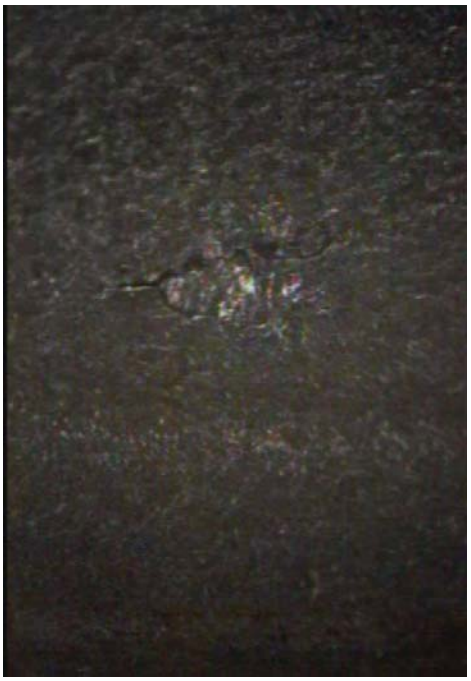


Figure 4.25 – Surface distress and micro-spalls observed on the remaining surface after 10,000,000 cycles using base oil BO1

2.2.2. Using base oil BO2

Observations on the dents

- After 500,000 cycles: the dents shoulder was perfectly visible and a small damaged area appeared.

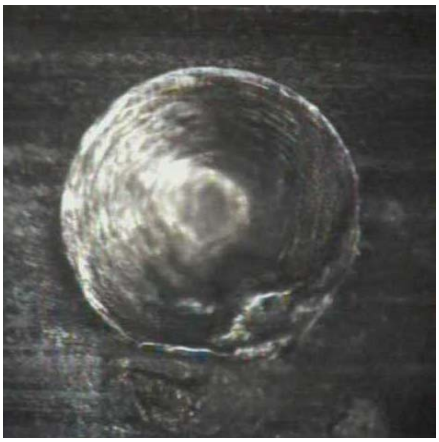


Figure 4.26 – Observations on a dent after 500,000 cycles using base oil BO2.

- After 1,000,000: a generalized damaged area around the dents was observed.

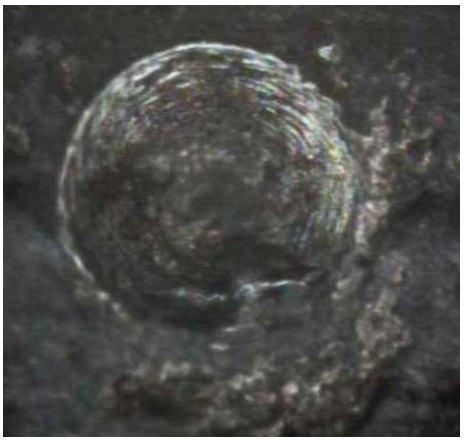


Figure 4.27 – Observations on a dent after 1,000,000 cycles using base oil BO2.

- Between 2,000,000 – 12,000,000: the generalized damaged area around the dents increased gradually and micro-spalls were observed next to the dents.

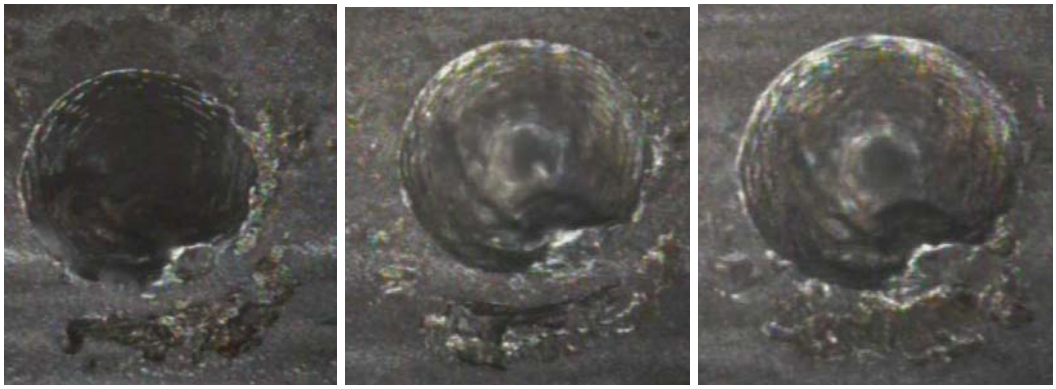


Figure 4.28 – Observations on the dents after 12,000,000 cycles using base oil BO2.

Observations on the remaining surface

- After 4,000,000 cycles it was observed that the surface roughness became smoother due to the gradual surface wear.
- After 12,000,000 cycles, generalized surface distress was observed on the remaining surface.

Figure 4.29 – Surface distress observed on the remaining surface after 12,000,000 cycles using base oil BO2

2.2.3. Using base oil BO3

Observations on the dents

- After 500,000 cycles: the dents shoulder disappeared and a damaged area appeared.

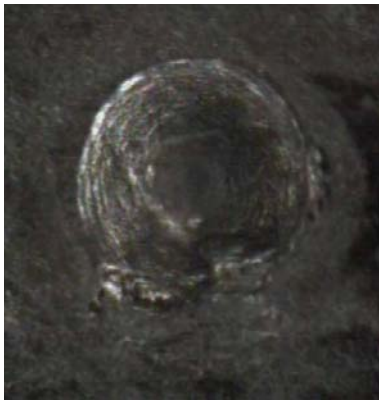


Figure 4.30 – Observations on a dent after 500,000 cycles using base oil BO3.

- Between 1,000,000 – 12,000,000 cycles: the generalized damaged area around the dents increased gradually and micro-spalls were observed next to the dents.

Figure 4.31 – Observations on the dents after 12,000,000 cycles using base oil BO3.

Observations on the remaining surface

- After 1,000,000 cycles it was observed that the surface roughness became smoother due to the gradual surface wear.

In Figures 4.32 and 4.33 the results of damage evolution for the different lubricants are shown. The photographs shown for each lubricant correspond to the most damaged dent of each disc.

Table 4.1 summarizes the damages found in the experimental tests as a function of the number of cycles.

Table 4.1 – Types of damages observed on the dents and remaining surface of the discs

		Observations on the dents		Observations on the surface	
		Micro-cracks around the dents	Spall around the dent	Surface distress	Natural dents
Lubricant	G1	3×10^6 cycles	4×10^6 cycles	—	3×10^6 cycles
	G2	1×10^6 cycles	2×10^6 cycles	—	—
	G3	5×10^6 cycles	—	—	—
	BO1	2×10^6 cycles	—	10×10^6 cycles	—
	BO2	1×10^6 cycles	—	12×10^6 cycles	—
	BO3	4×10^6 cycles	—	—	—
		Number of fatigue cycles			

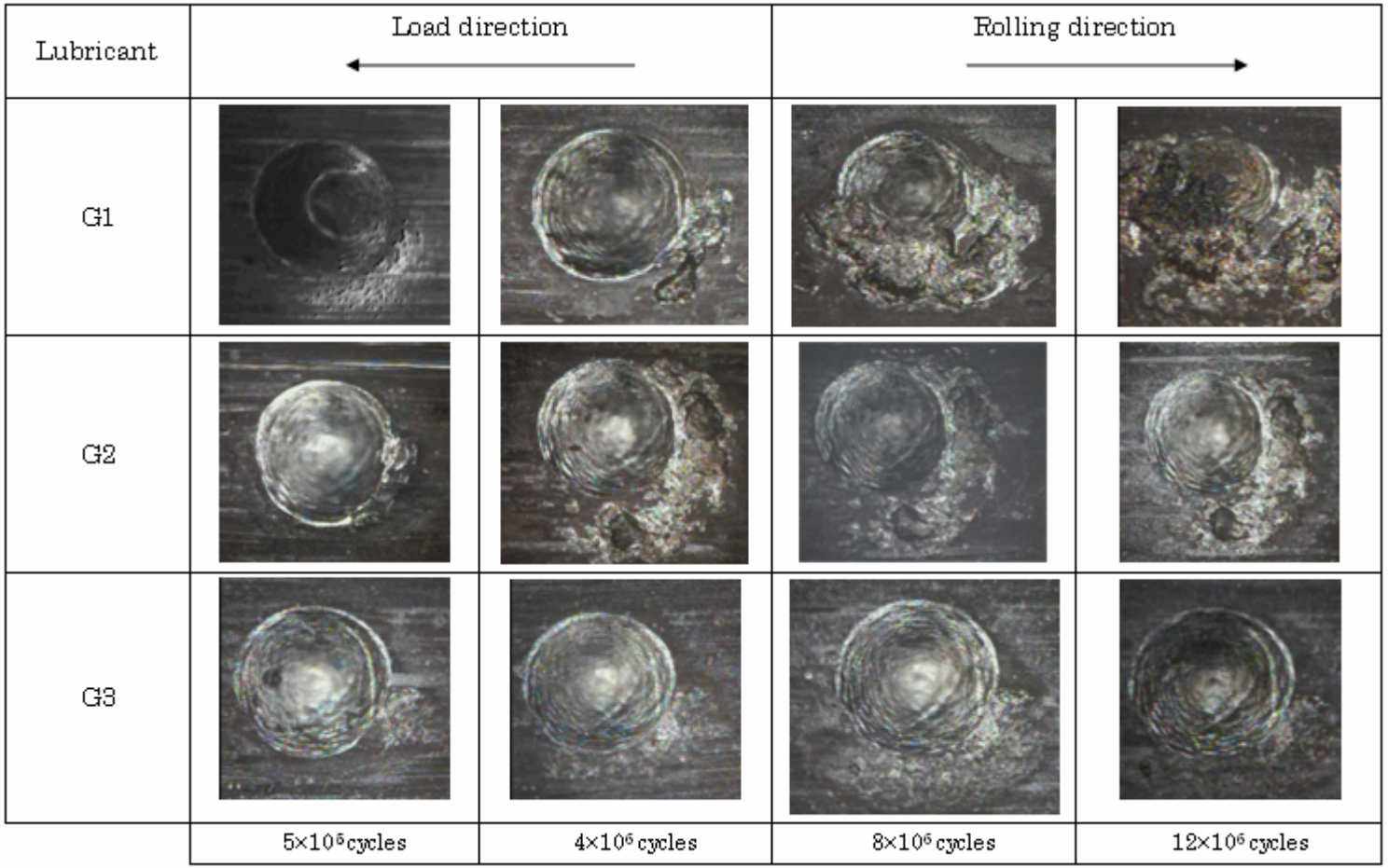


Figure 4.32 – Aspects of dent fatigue damage area evolution for the greases tested.

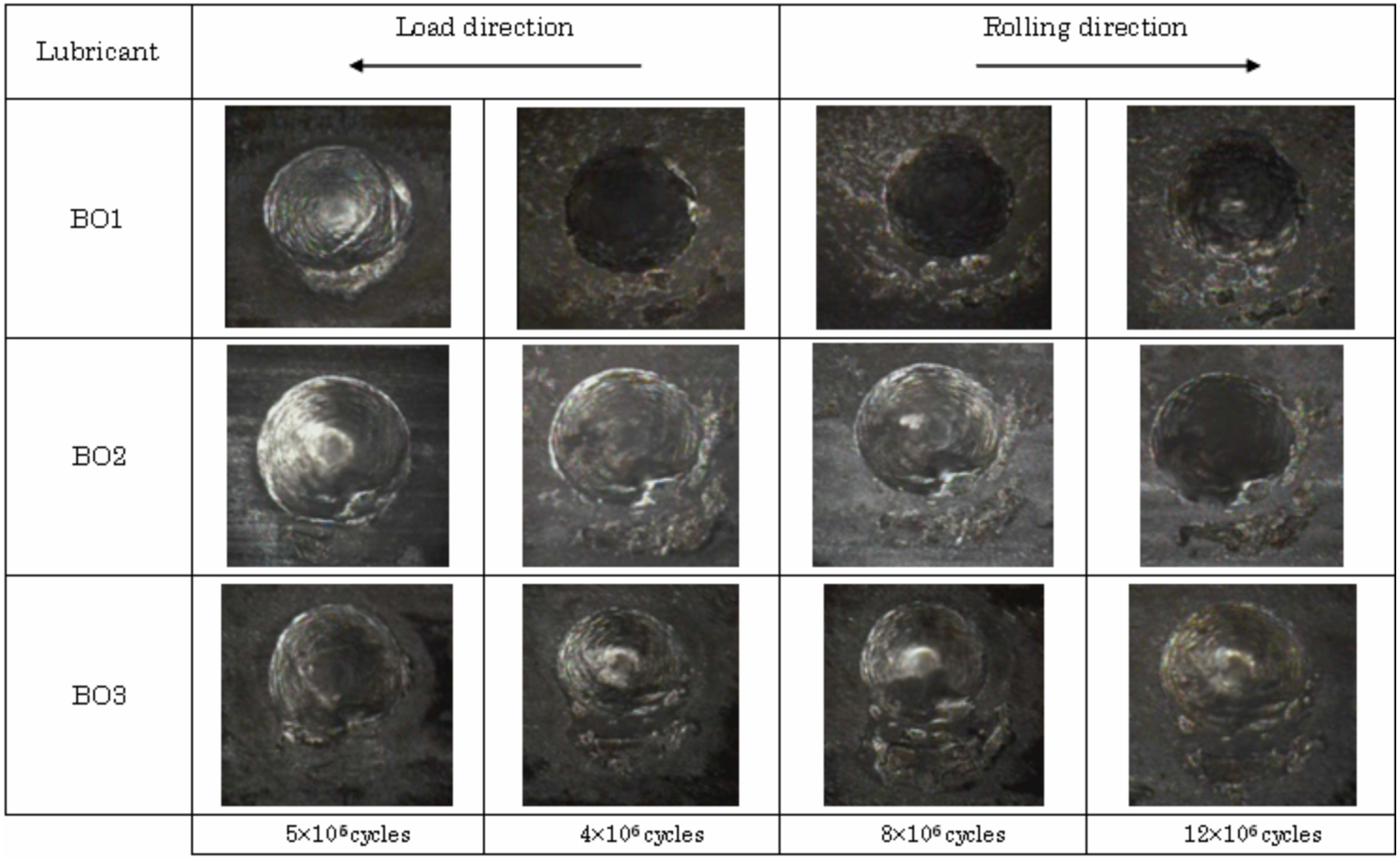


Figure 4.33 – Aspects of dent fatigue damage area evolution for the base-oils tested.

2.3. Artificial dents damaged area growth

The images exposed above revealed that experimental tests with different lubricants resulted in damaged areas around the dents with different sizes. Consequently, it was found to be interesting calculating this area and compare it between all the tests.

When comparing the three dents of the same disc for a given number of cycles it was found that not all the three dents printed in the spherical surface suffered the same damage. This was due to the fact that not all the dents were perfectly printed in the center of the contact, and consequently were not subjected to the same Hertzian pressure. Given this fact, it was decided that in order to compare the damaged areas around the dents for the different tests, it had to be made between the dents of each test that suffered more damaged, that is those that were more accurately printed in the center of the contact and that were consequently subjected to the maximum Hertzian pressure.

To calculate the damaged area growth around the dents, the images taken from the artificial dents throughout the tests were analyzed using the commercial *AutoCAD* drawing software. The method to calculate the damaged area consists in copying the pictures to the program's interface and then contouring the damaged area. After the contour around is drawn the software displays the area (Figure 4.34).

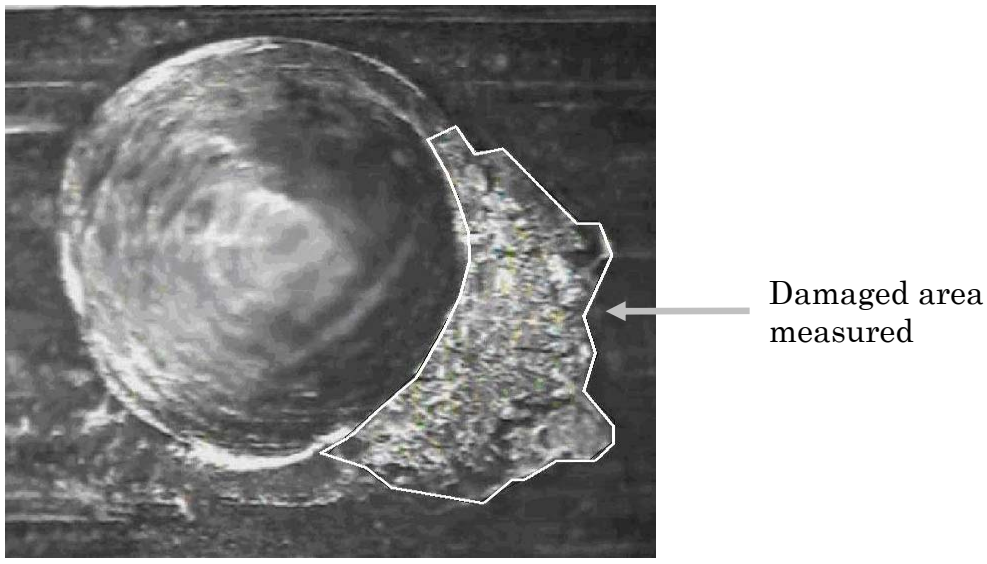


Figure 4.34 – Method used to evaluate the damaged area around the artificial dents.

This method is simple and quick to use but it can be affected by several errors especially when obtaining the pictures using the video microscope. The most

significant ones are related with the manual focus that might not be the same for all the pictures and the incorrect position of the lenses in relation to the dents, which might deform the image and show a dent that has not a circular shape but an oval one.

The damaged area calculation results are presented in Figure 4.35 where a comparison for all the lubricants is shown. Figures 4.36 and 4.37 present a comparison of the damaged area evolution for the greases and base oils respectively. To have a better perception of how the damaged area evolutes for each pair grease / base oil, it was found to be pertinent presenting the results for each pair (Figures 4.38 to 4.40).

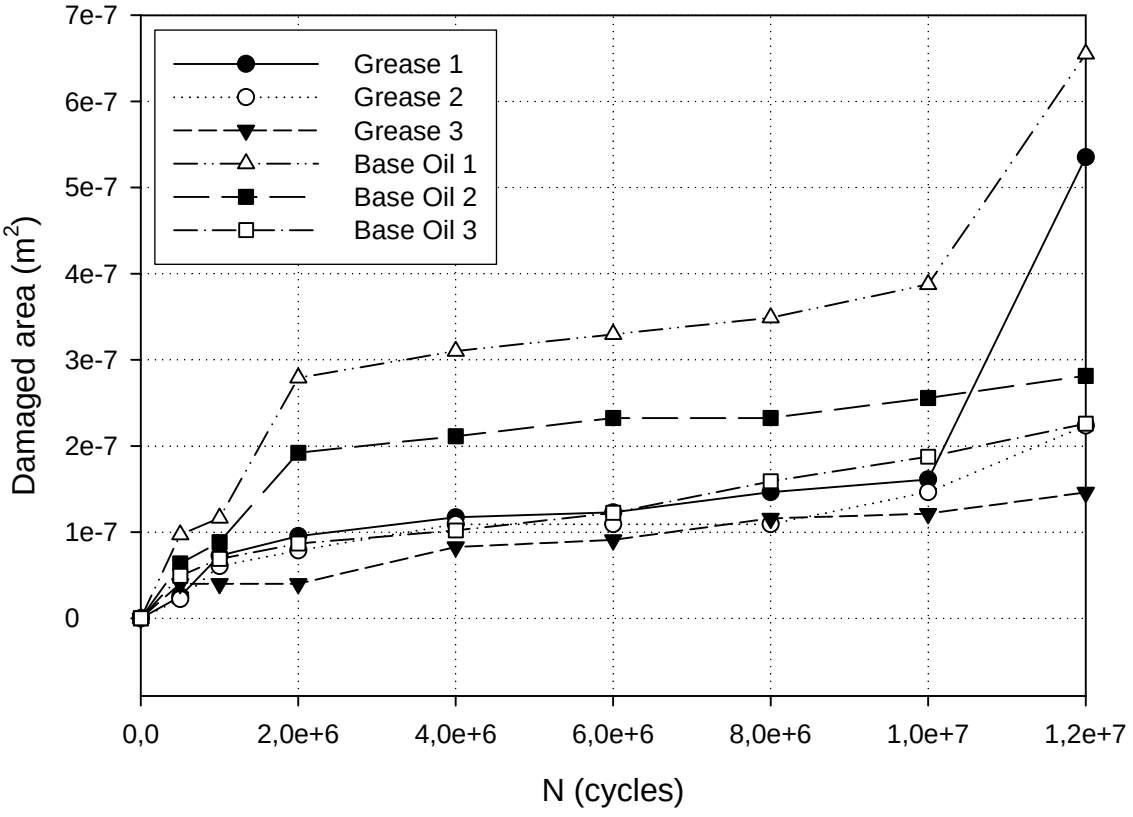


Figure 4.35 – Evolution of the damaged area around the dents for all the lubricants tested.

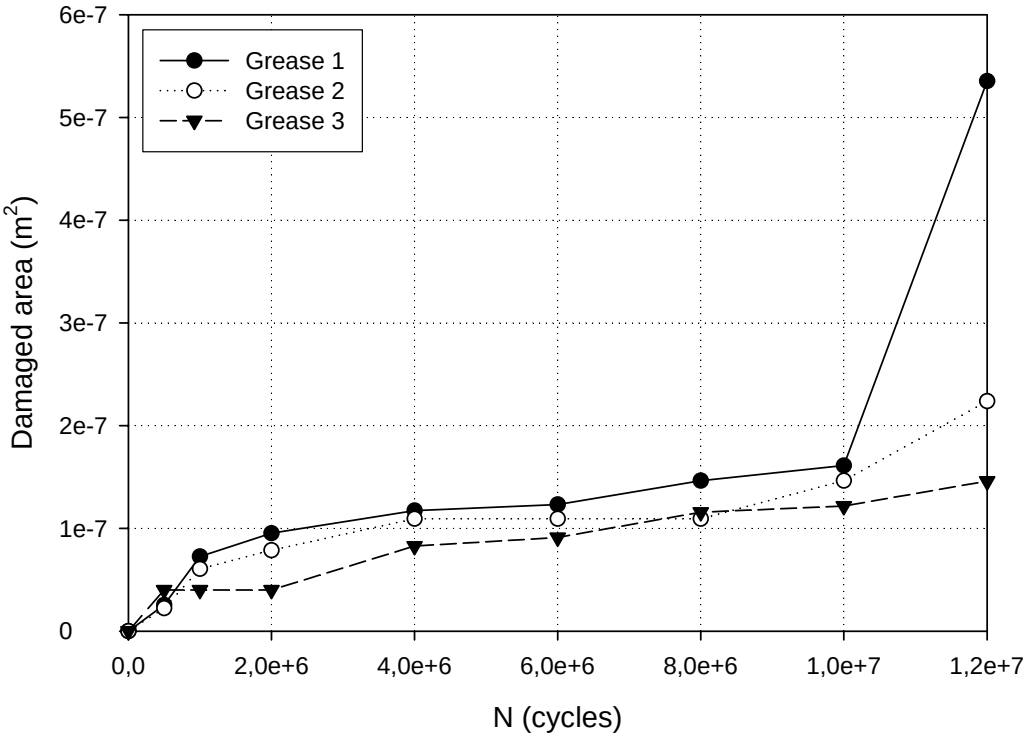


Figure 4.36 – Evolution of the damaged area around the dents for the greases.

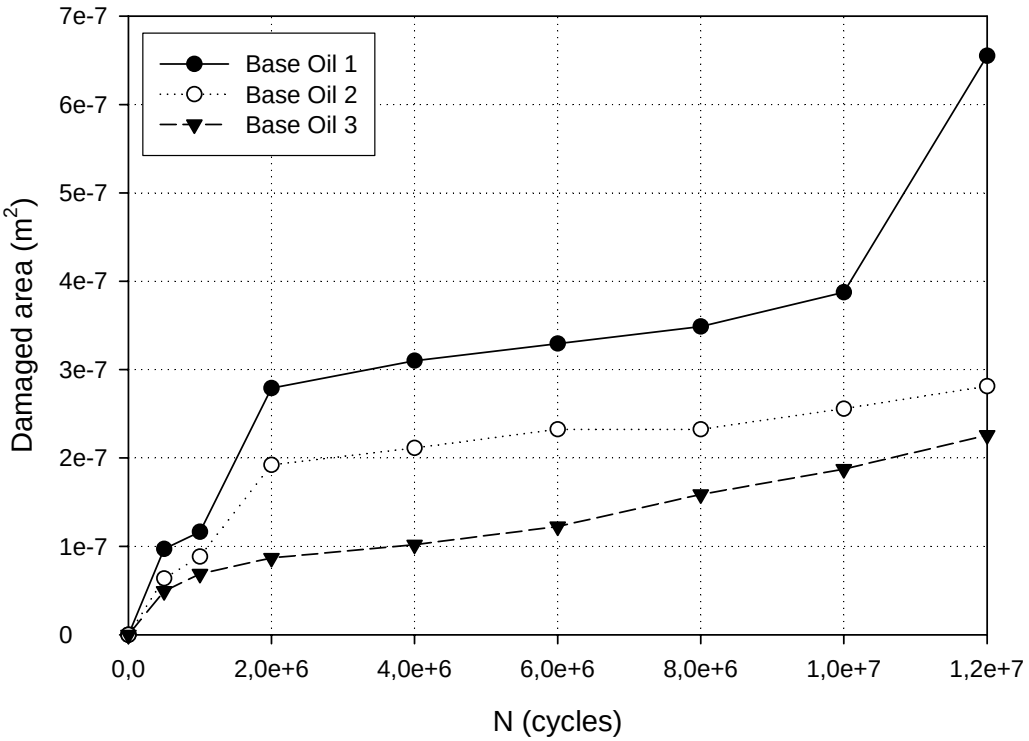


Figure 4.37 – Evolution of the damaged area around the dents for the base oils.

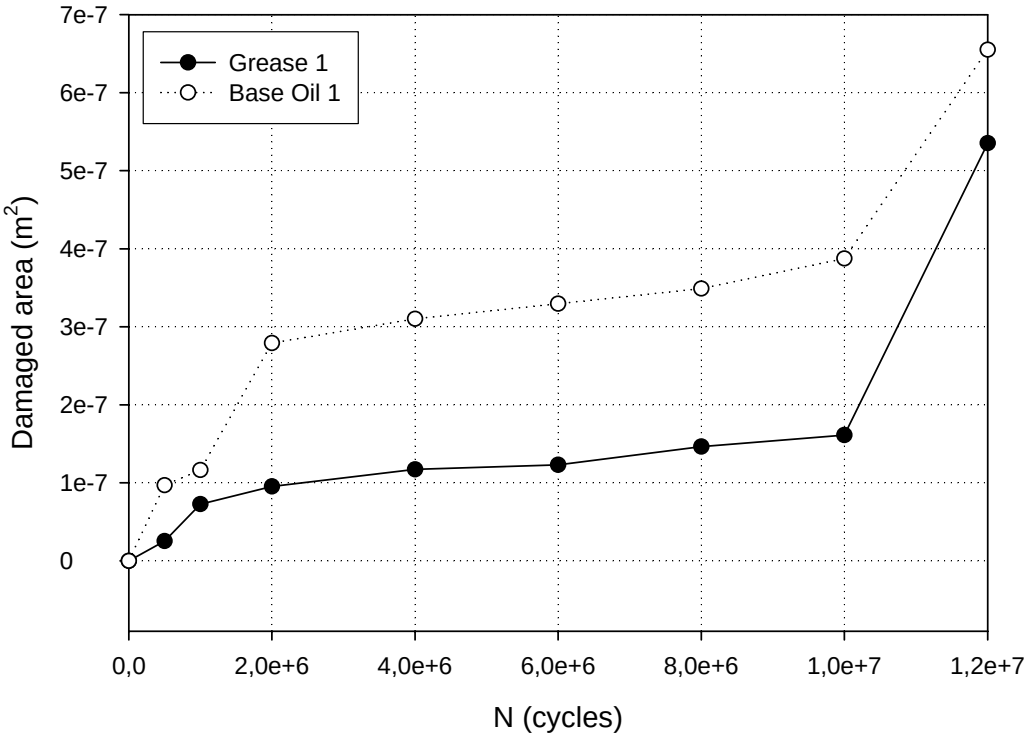


Figure 4.38 – Evolution of the damaged area around the dents for the pair G1/BO1.

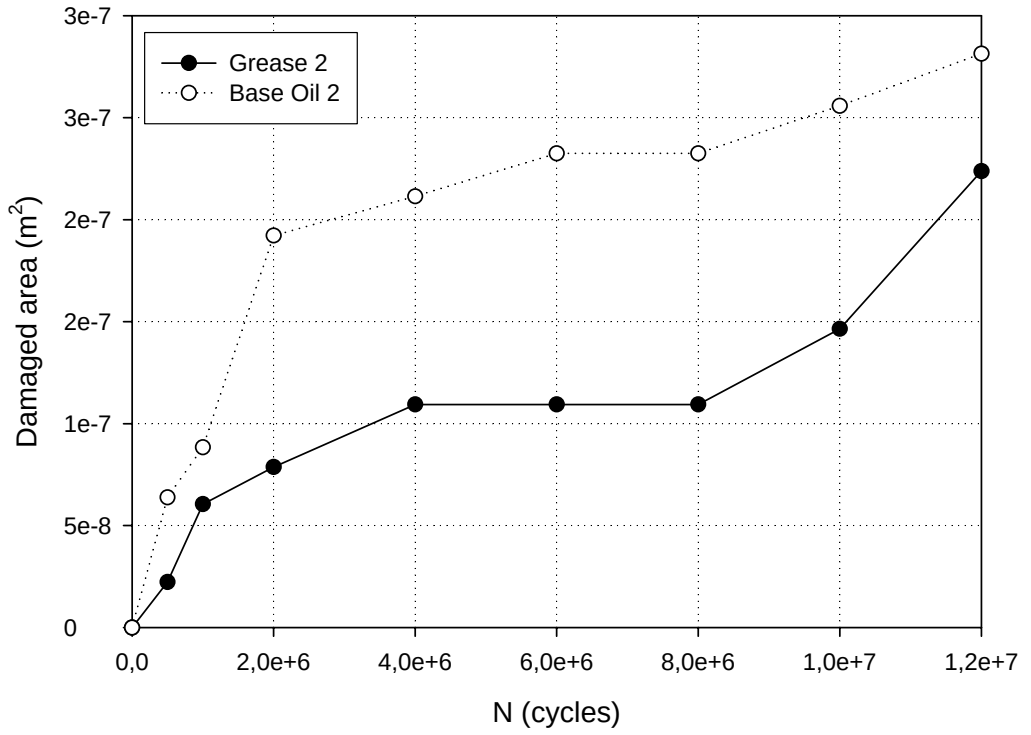


Figure 4.39 – Evolution of the damaged area around the dents for the pair G2/BO2.

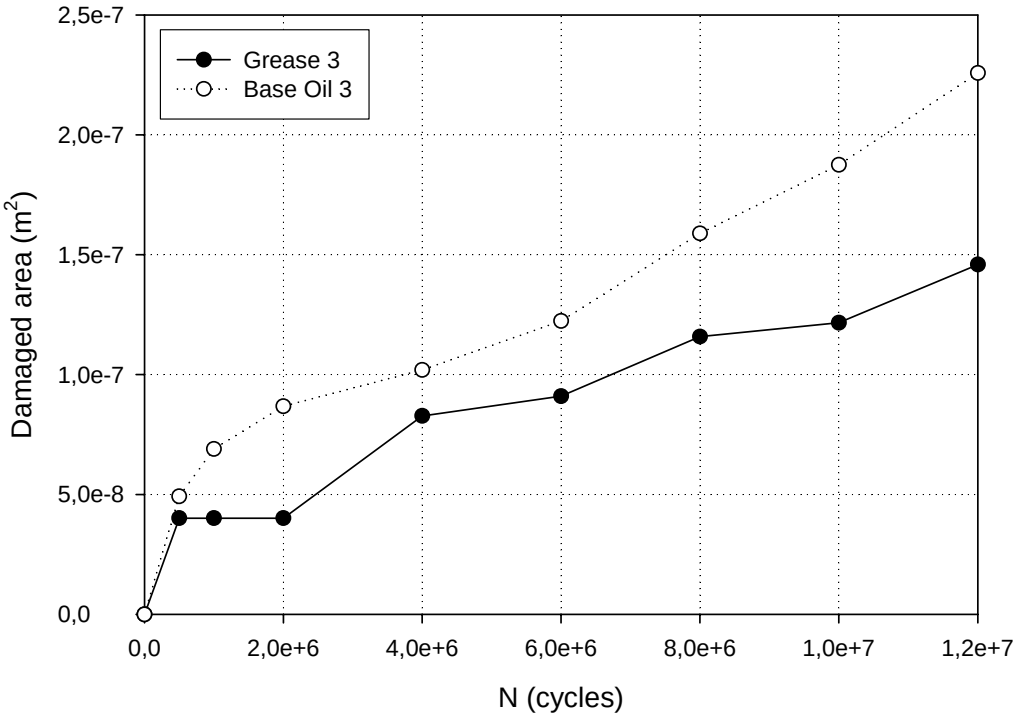


Figure 4.40 – Evolution of the damaged area around the dents for the pair G3/BO3.

2.4. Volume wear calculation

As it can be seen in Figure 4.41, where the initial and final configurations of a dent are shown, there is an amount of material that is removed from the surfaces throughout the experimental tests. Given this fact, and because lubricants influence surface wear, the amount of material removed (wear volume) was calculated and the results obtained were compared for the greases and base oils experimentally tested.

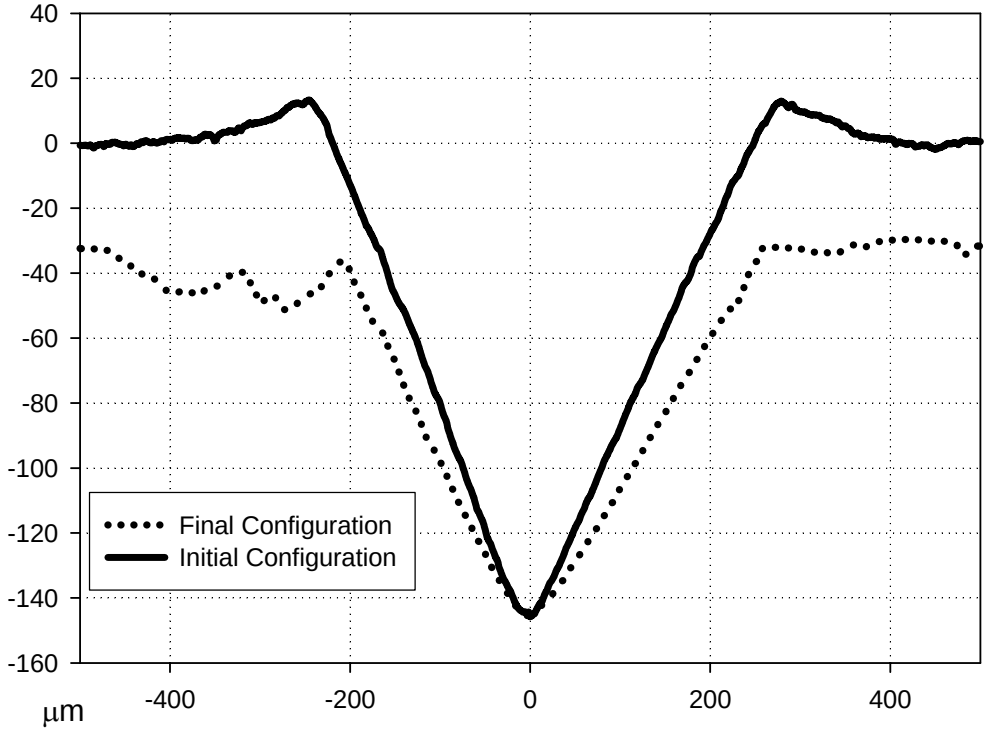


Figure 4.41 – Comparison between the initial and final configurations of a dent revealing the amount of material removed from the surfaces.

To calculate the wear volume, the artificial dents were used as marker. In Figure 4.42, a scheme of the contacting zone that suffers progressive wear is shown. The wear volume V_w can be determined from the value of the wear area A_w by the simple relation:

$$V_w = 2\pi A_w \tag{4.1}$$

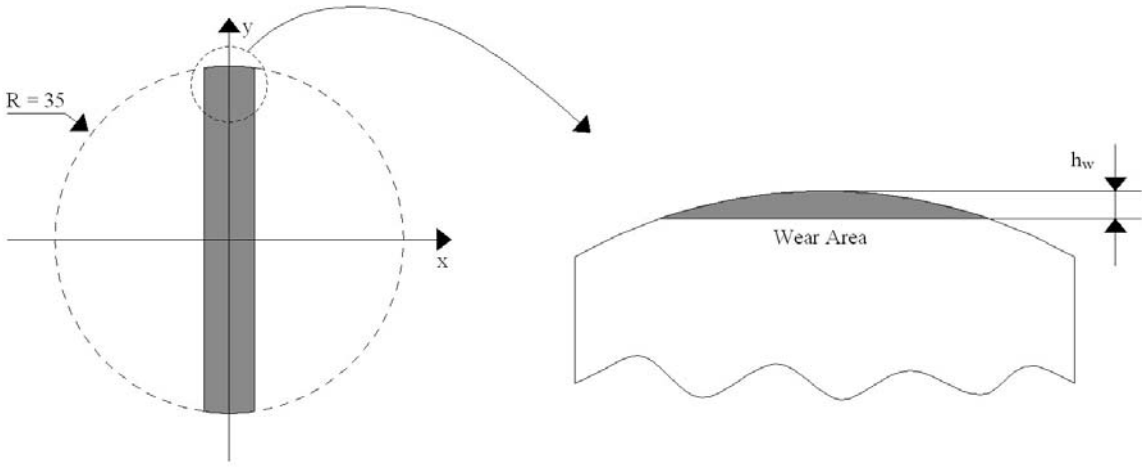


Figure 4.42 – Scheme of the contacting zone that suffers progressive wear.

The value of A_w depends on the wear height h_w which can be related to the dent radius r_d that diminishes during the experimental tests, as shown schematically in Figure 4.43.

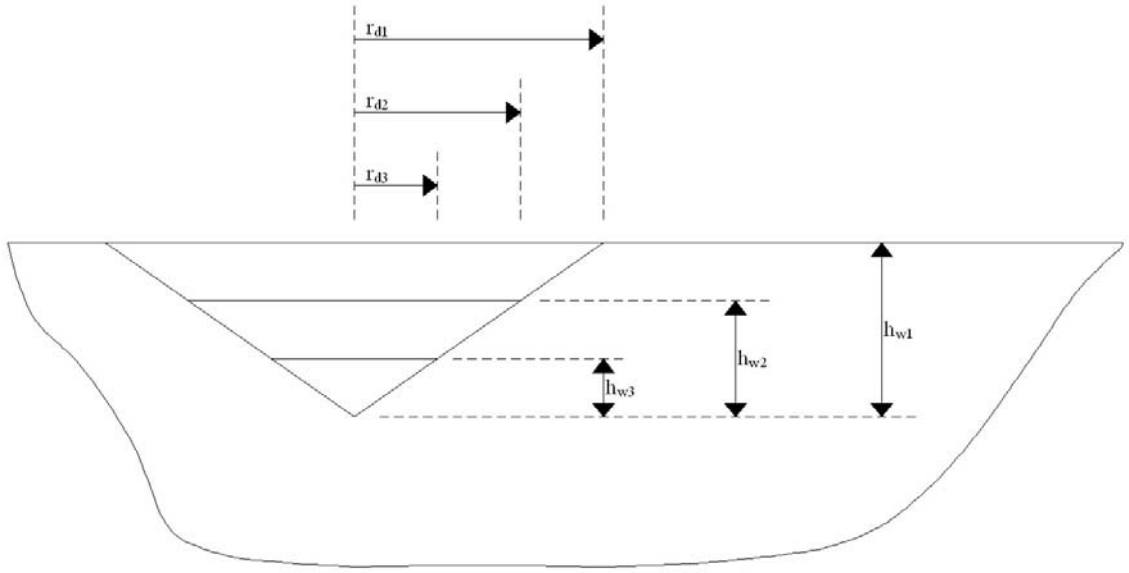


Figure 4.43 – Relation between the wear height h_w and the dent radius r_d.

The values of the wear height h_w are related to the dent radius r_d by the following equation:

$$\frac{r_{d1}}{h_{w1}} = \frac{r_{d2}}{h_{w2}} = \cdots = \frac{r_{dn}}{h_{wn}} \tag{4.2}$$

The values of h_{w1} and r_{d1} are known at the beginning. For every stop in the experimental tests the values of r_{dn} were determined (using the same method used for the damaged area calculation) and consequently the values of h_{wn}.

The wear area A_w (Figure 4.44) is given by:

$$A_w = 2\left[\int_{l_1}^{l_2} \sqrt{R^2 - x^2}\,dx - \int_{l_1}^{l_2} (R - h_w)dx\right] \tag{4.3}$$

where

$$l_1 = 0$$

$$l_2 = \sqrt{2Rh_w - h_w^2} \tag{4.4}$$

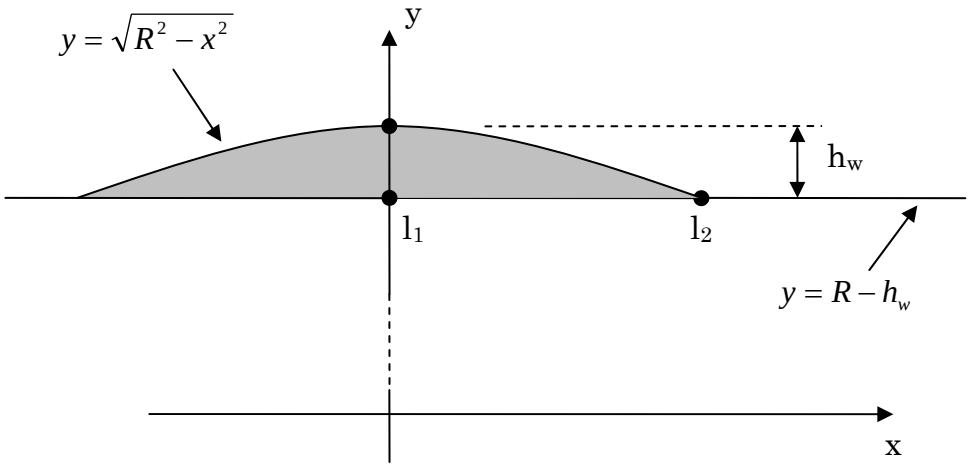


Figure 4.44 – Area wear A_w calculation.

The results for the wear volume calculations for the different lubricants tested are given in Table 4.2 and Figure 4.45. Figures 4.46 and 4.47 present a comparison of the wear volume for the greases and base oils respectively. To have a better perception of how the wear volume evolutes for each pair grease / base

oil, it was found to be pertinent presenting the results for each pair (Figures 4.48 to 4.50).

Table 4.2 – Wear volume for the various lubricants tested.

Cycles	Wear Volume (m^3)					
	Lubricant					
	G1	G2	G3	BO1	BO2	BO3
5×10^5	1,14E-07	7,34E-08	9,29E-09	4,75E-08	1,7E-07	9,29E-09
1×10^6	1,35E-07	1,03E-07	9,29E-09	2,91E-07	2,91E-07	9,29E-09
2×10^6	2,08E-07	1,35E-07	9,29E-09	2,91E-07	2,91E-07	9,29E-09
4×10^6	2,48E-07	2,08E-07	9,29E-09	2,91E-07	2,91E-07	7,34E-08
6×10^6	3,34E-07	2,48E-07	9,29E-09	3,34E-07	2,91E-07	1,02E-07
8×10^6	3,34E-07	2,48E-07	9,29E-09	3,34E-07	2,91E-07	1,02E-07
10×10^6	3,81E-07	2,48E-07	9,29E-09	3,81E-07	3,81E-07	1,02E-07
12×10^6	7,02E-07	3,34E-07	2,57E-08	3,81E-07	3,81E-07	1,02E-07

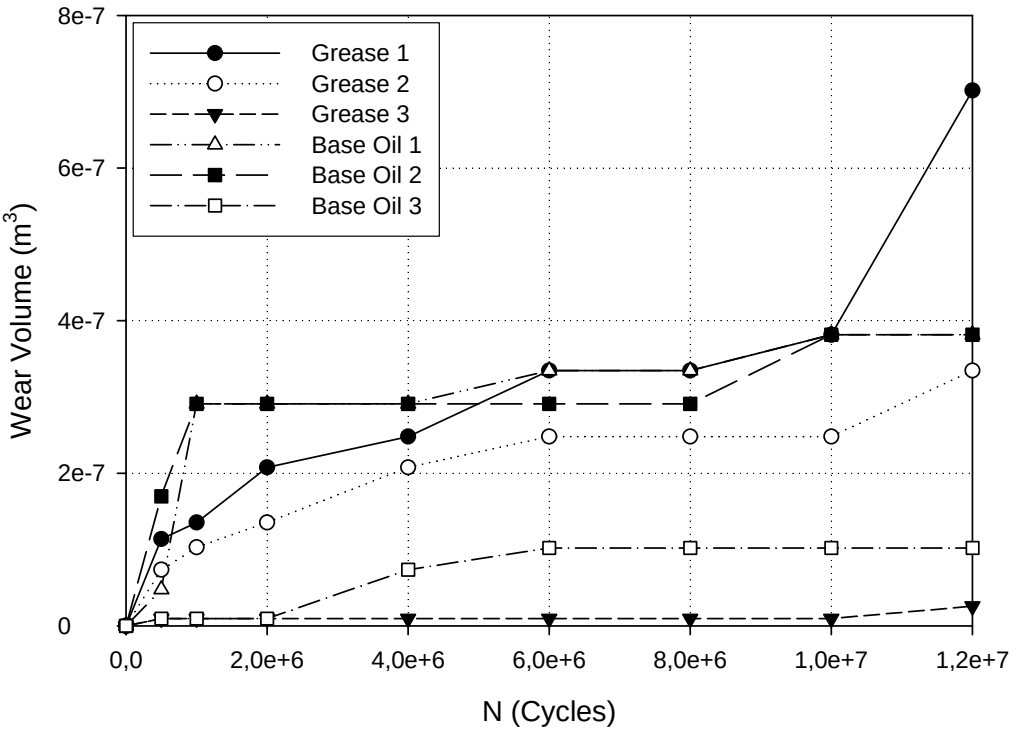


Figure 4.45 – Evolution of the wear volume for all the lubricants tested.

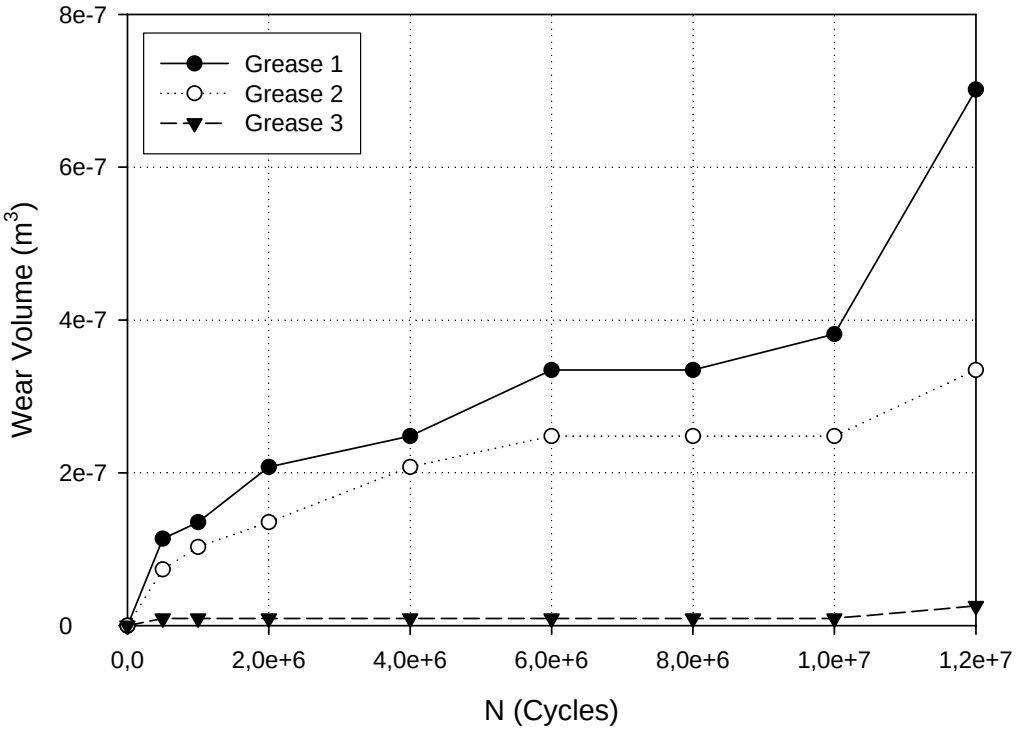


Figure 4.46 – Evolution of the wear volume for the greases.

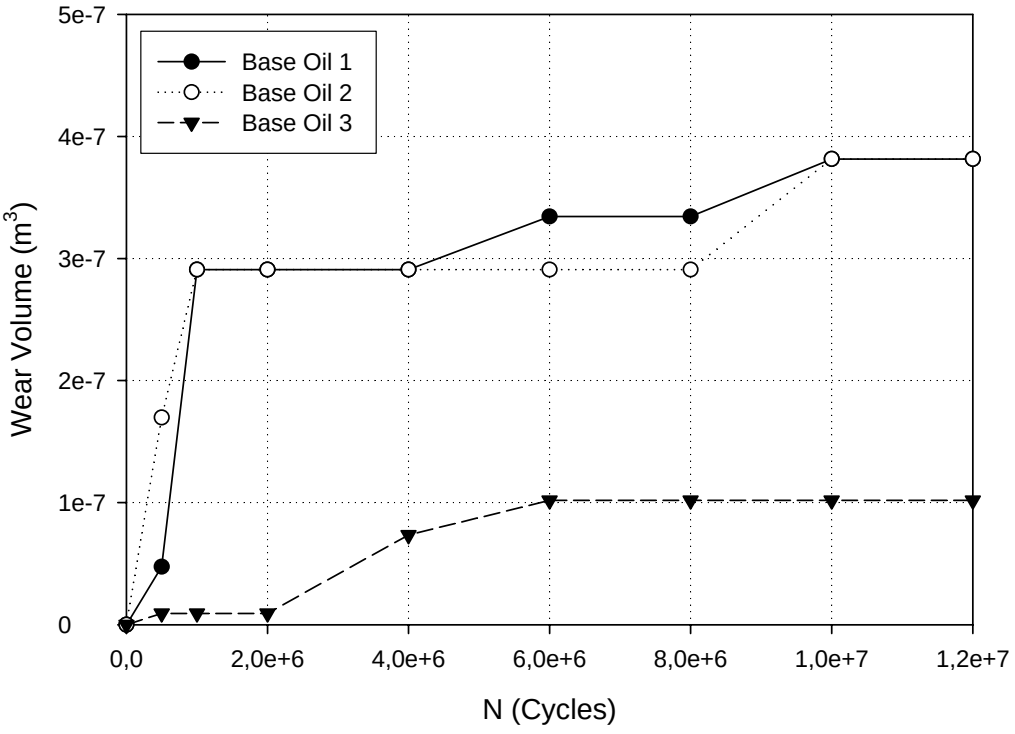


Figure 4.47 – Evolution of the wear volume for the base oils.

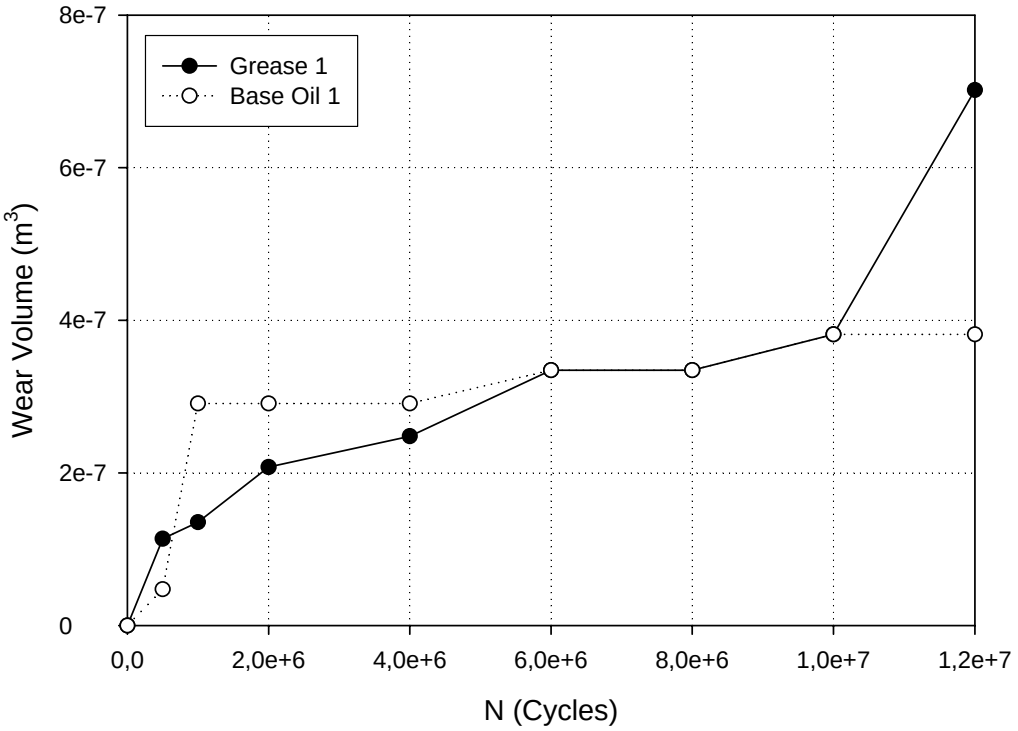


Figure 4.48 – Evolution of the wear volume for the pair G1/BO1.

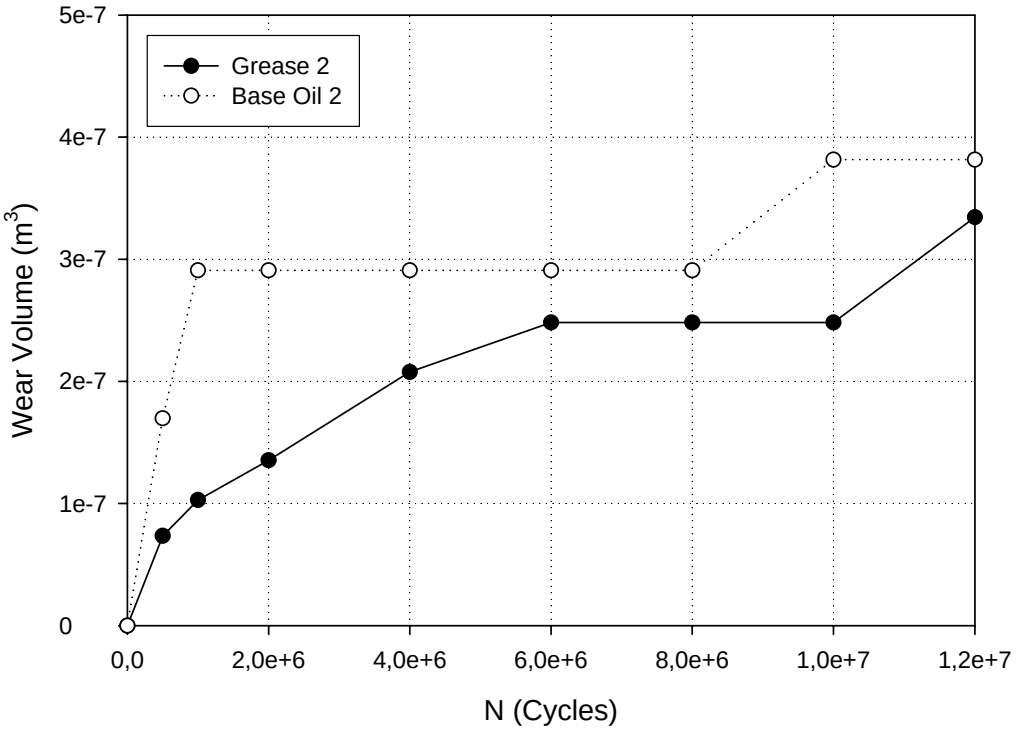


Figure 4.49 – Evolution of the wear volume for the pair G2/BO2.

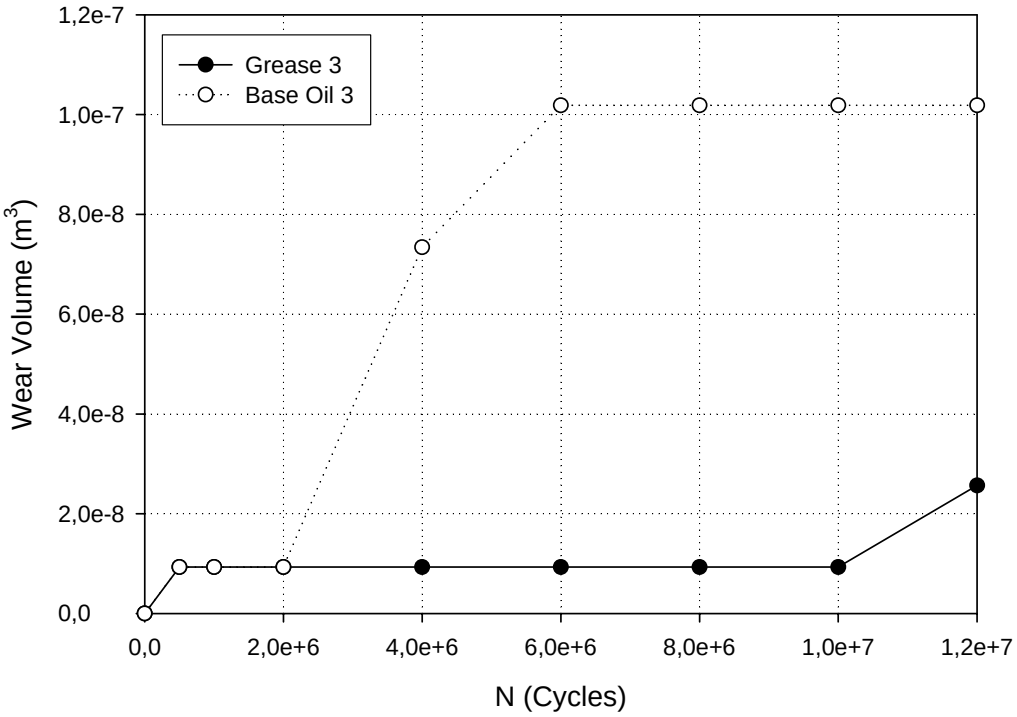


Figure 4.50 – Evolution of the wear volume for the pair G3/BO3.

3. Sub-surface observations

To complete the surface observations and to analyze the crack initiation sites and the crack propagation directions, cross-sections of the dents of two discs were performed and analyzed using a scanning electron microscopy (SEM) at CEMUP. These discs correspond to those that were used for the tests using respectively the lubricants G3 and BO1. They were chosen because in these two no surface spall was still visible after the test, suggesting that cracks propagation could be observed (contrarily to what happened in the tests using the greases G1 and G2).

3.1. Cracks propagation

It was observed that cracks propagate in the zone under the dents shoulder of the artificial dents (Figure 4.51).

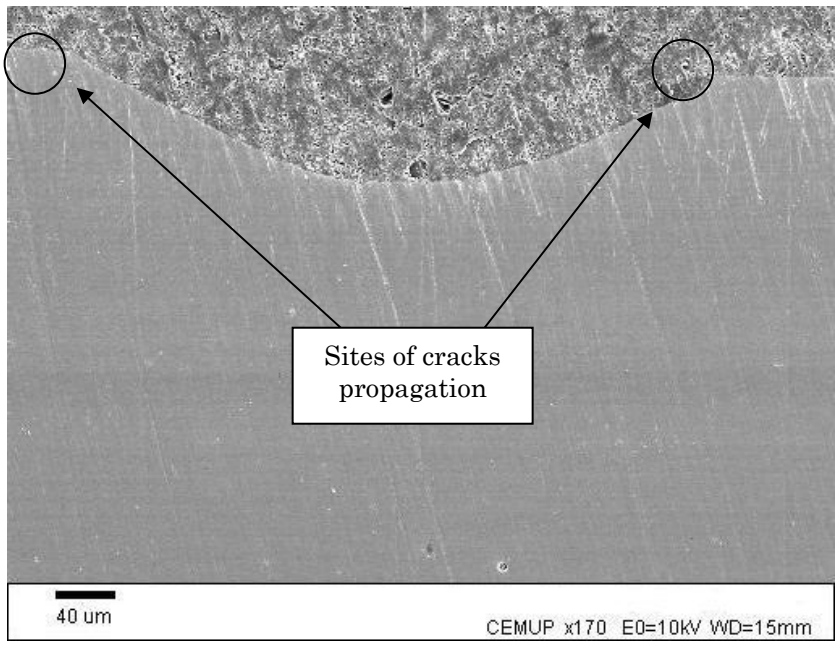


Figure 4.51 – Sites were cracks were observed for the test using BO1.

Figures 4.52 to 4.54 show propagated cracks under the dent shoulder of an artificial dent.

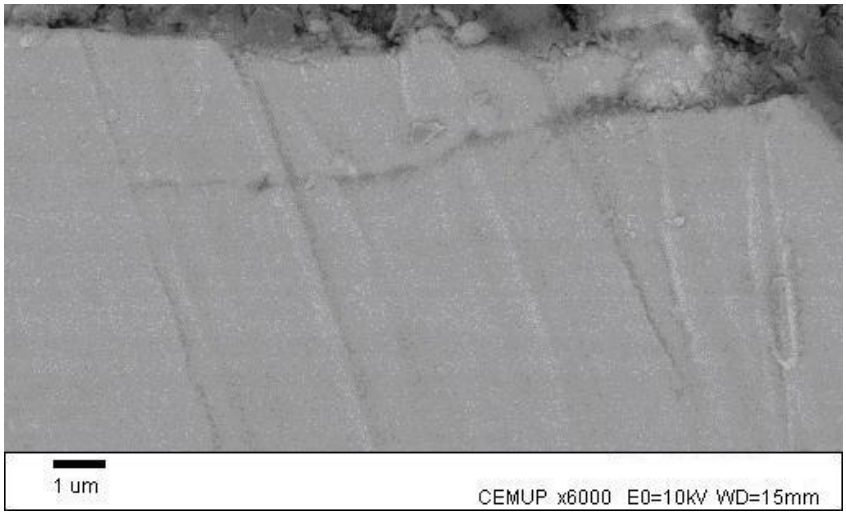


Figure 4.52 – Crack propagation under the dent shoulder for the test using the BO1.

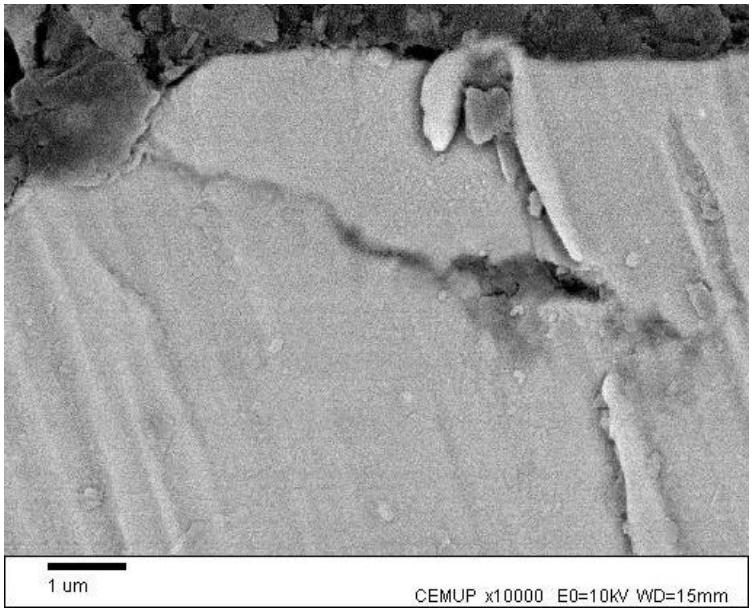


Figure 4.53 – Crack propagation under the dent shoulder for the test using the BO1.

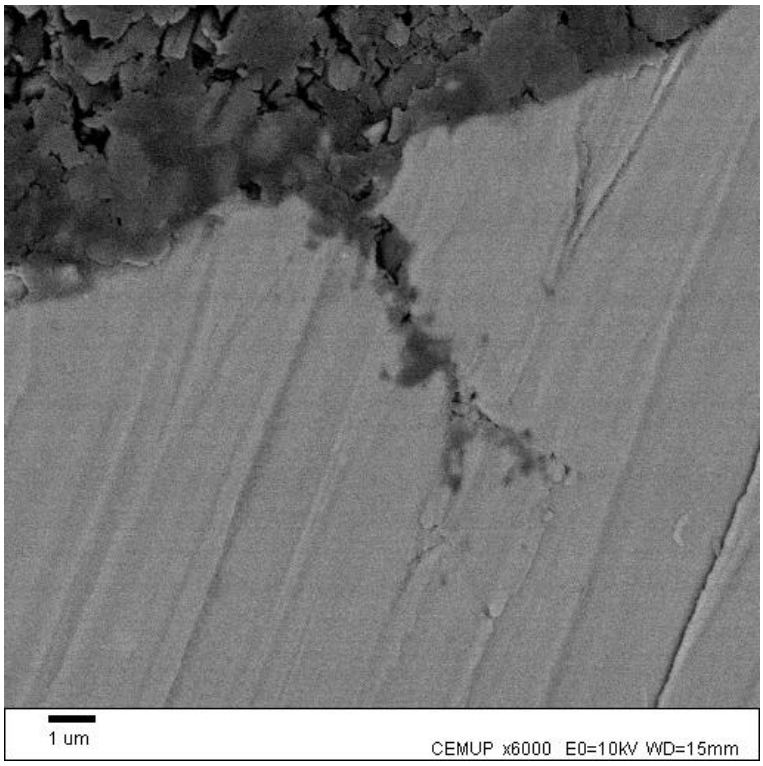
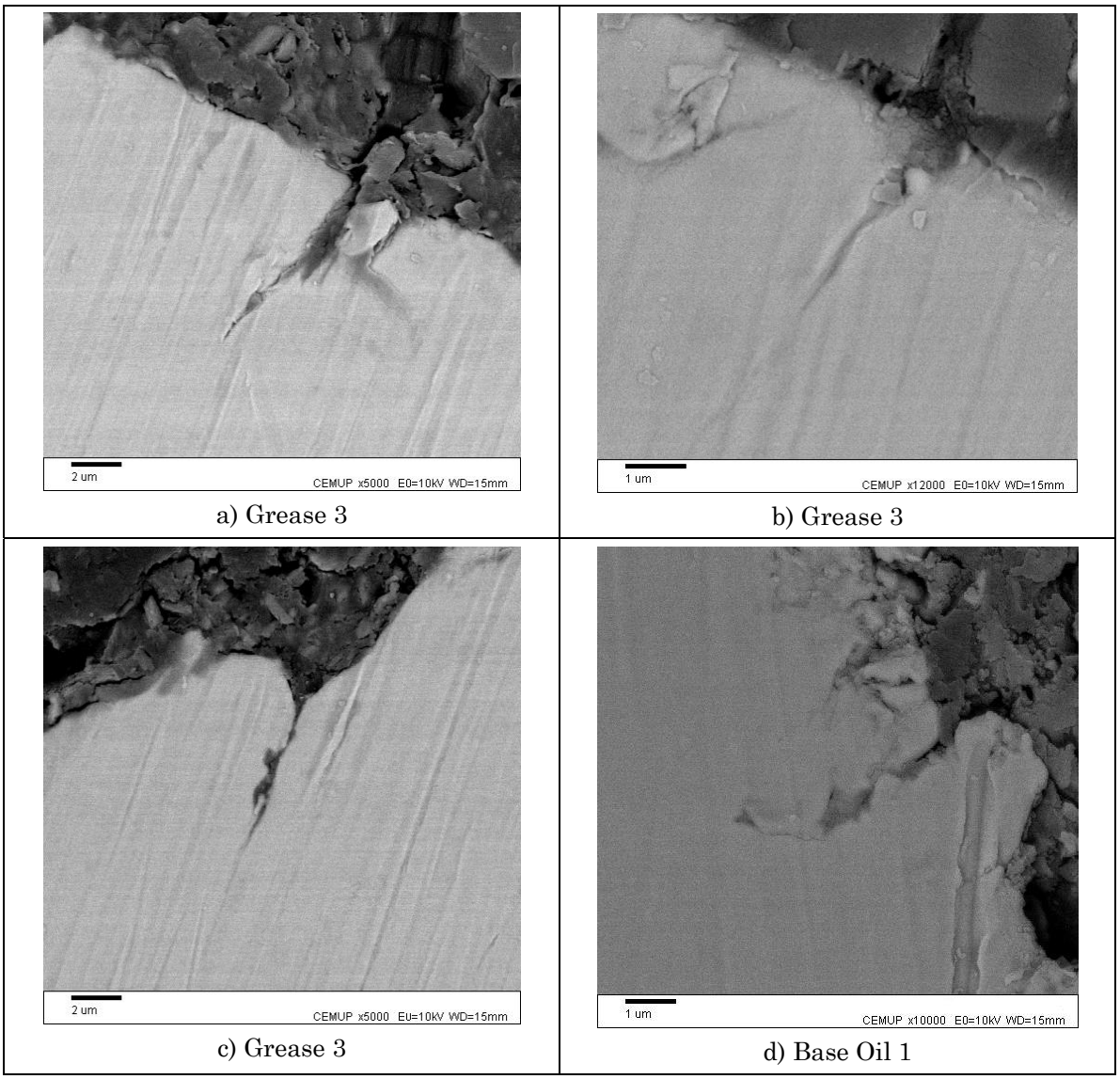


Figure 4.54 – Crack propagation under the dent shoulder for the test using the G3.

Cracks inside the dents were also observed, as it is shown in Figure 4.55. The origin of these cracks is probably related to the indent process rather than to a fatigue process although they might have propagated during the experimental tests. During the indent process, due to inhomogeneity of the material, some cracks might appear when the indenter penetrates the surface.



a) Grease 3 b) Grease 3

c) Grease 3 d) Base Oil 1

Figure 4.55 – Cracks observed inside the dents for tests using BO1 and G3.

3.2. Spall craters

Experimental tests using grease G1 and G2 ended with spall craters ahead the dents. The profile of such spalls was obtained using a profilometer from CETRIB. The results are shown in Figure 4.56 were the propagation angles are also displayed.

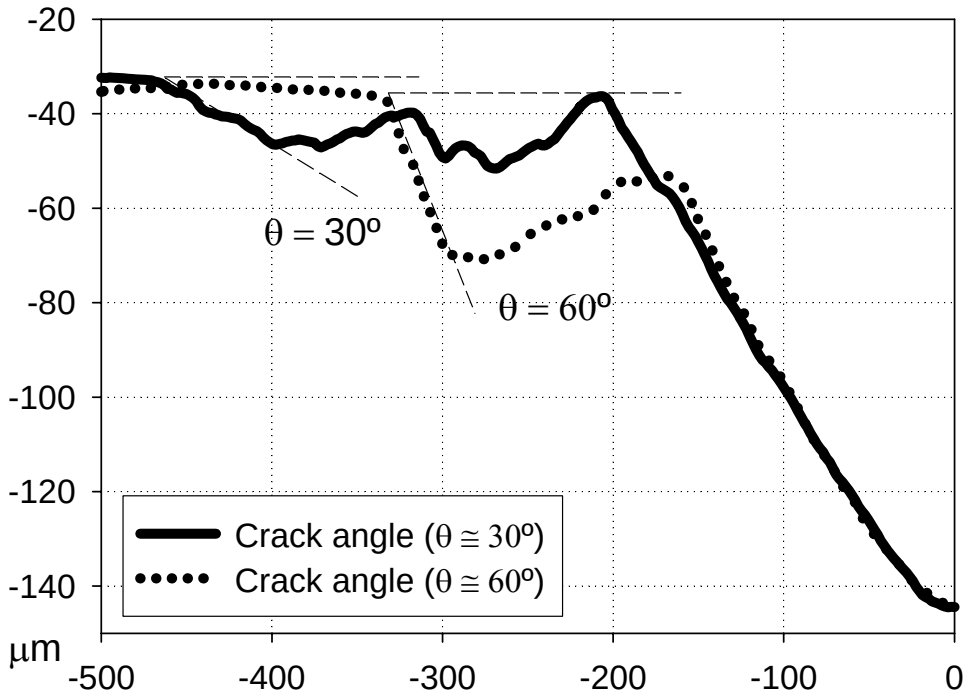


Figure 4.56 – Profile of two spall craters with the inclination angles of crack propagation.

4. Results discussion

Location of surface damage on the dents

The location of the damaged area for each dent tested was almost always ahead the dent along the direction of load movement (Figure 4.57). This is coherent under pure rolling conditions with the results obtained by other authors [46,50,51]. The reason for these observations can be found in transient elastohydrodynamic lubrication analysis, which predicts a slightly higher pressure peak ahead of the dent, along the load movement direction under pure rolling [45] as it was explained in Chapter I, Section 5.

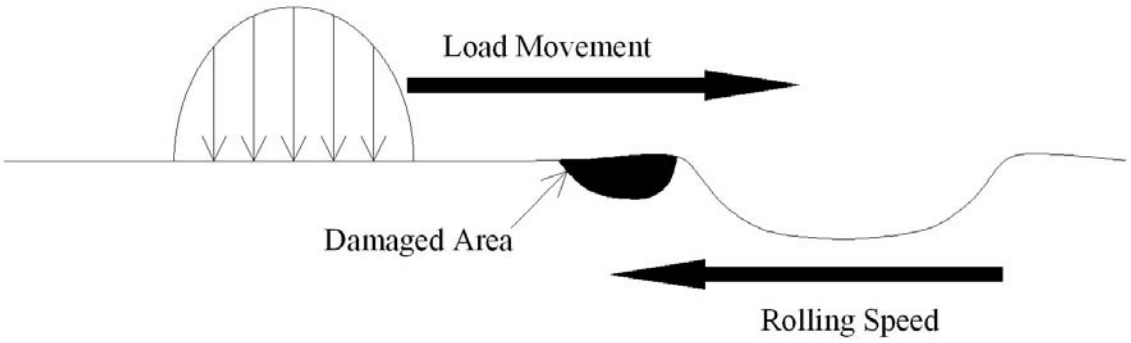


Figure 4.57 – Location of the fatigue damaged area.

In some cases, it was found that the damaged area was not ahead the dent, but micro-cracks and incipient spalls were observed on one of the sides of the dents. This was probably due to a bad accuracy in printing the dents in the centre of the discs.

Dents damage mechanism

The dents damage mechanism that was found during the experimental tests can be generally summarised as follow:

- Whatever the lubricant may be, the raised shoulder is flattened by the high pressure spikes generated when the contacting surface move on the dent edge. The shoulder height decreases very quickly and becomes quite stable long after.
- Due to the accumulated plastic strain resulting from the high pressure

spikes and the high stress concentration beneath the edge dent, cracks are initiated on or close to the surface at the dent edge.

- These cracks might propagate into a large spall depending on the lubricant used.

Sub-surface observations revealed that cracks propagated beneath the dent shoulder which is certainly due to the high shear stresses induced by the shoulder subsidence.

It was also observed that in some cases spall craters were formed with a depth and size that depended on the initiation angle. The bigger was the initiation the angle, the deeper and larger was the final crack. Crack initiation angles between 30° and 60° were found.

Damaged area evolution around the dents

Concerning the evolution of the damaged area growth it can be observed that it is more or less similar for all the tests. Generally, three phases can be distinguished:

1. A first phase (until $\approx 2\times10^6$ cycles) characterized by a fast initial growth of the damaged area due to the high plastic deformation of the dents shoulder in which micro-cracks appear;
2. A long and intermediate phase (between 2×10^6 and 10×10^6 cycles) characterized by a slow growth of the damaged area;
3. A final phase in which micro-cracks that appeared around the artificial dents grow to form a spall.

The damaged area around the dents was found to be always larger for the tests using base oils than for those using greases. Nevertheless some errors might exist influencing these results. Among those is the lack of accuracy in printing the dents in the centre of the discs (surface observations suggest this because the damaged area was not always found ahead the dents) and the errors already referred in Section 2.3. Furthermore, this analysis has another limitation because the damaged area has different natures. Two equal damaged areas might correspond in one case to a spall crater and in another to a surface exhibiting micro-cracks.

Evolution of the surfaces wear

Concerning the evolution of the surface wear it was found that it was not similar for all the tests. Even though two phases could be clearly distinguished:

1. A first phase, that can be shorter or longer depending on the lubricant used, characterized by a greater initial wear due the more frequent contact between the asperities;
2. A second phase, characterized by a progressive and slow wear in which the surface is flattened and the contact between the asperities is less frequent;

It was found that for each pair grease/base oil the wear volume was always higher for the tests using the base oils. An exception was found for the pair G1/BO1 for which the wear volume calculated at the end of the test (12×10^6 cycles) was higher for the contact lubricated with the grease. A reason that can explain this is that the amount of grease available decreased considerably and the contact became more starved which conduced to a smaller value of λ and consequently to a boundary film lubrication regime.

The surface wear volume calculation depends on the dent radius that was determined using same method described to calculate the damaged area around the dents. Consequently, also these results of wear volume might affected by the same errors referred in Section 2.3.

Lubricant film thickness influence

The influence of the lubricant film thickness is expressed by film thickness parameter λ which defines the lubrication regime. The experimental results obtained by Cheng *et al.* [23] have revealed that, when λ is smaller than 0.3, cracks are initiated everywhere on the surface and, when λ is greater than 0.5, no surface cracks could be found, except at the region near the surface defects. This is consistent with the results obtained; for all the tests λ is considerably greater than 0.5 and cracks were initiated only around the artificial dents. An exception was found for the test using grease G1, in which cracks were found on the remaining surface. Even though, the reason for this is not related with the value of λ but rather to the existence of metallic particles (probably released from the dents) that entered the contact. Maybe this can suggest a difficulty for grease G1 in expelling metallic particles although nothing clearly supports this hypothesis.

The value of λ also has an important influence in the wear suffered by the contacting surfaces and can be used as a mean to justify the results obtained. Tests using base oils had smaller values of λ than those using greases. In the case of base oils BO1 and BO2 the value of λ expressed a mixed film lubrication

regime. That also justifies the reason why surface distress was found in the surface of the discs using these two base oils (at 10×10^6 cycles using BO1 for λ = 1.184 and at 12×10^6 cycles using BO2 for λ = 1.321).

Lubricants composition influence

Comparing the results using greases it could be observed that tests using grease G3 (8.7 % soap concentration, 530.873 mm^2/s of base oil viscosity at 40 ºC and EP additive) suffered irrelevant damage around the dents. On the other hand, tests using grease G1 (9 % soap concentration, 100.000 mm^2/s of base oil viscosity at 40 ºC) and G2 (12 % soap concentration, 124.495 mm^2/s of base oil viscosity at 40 ºC) suffered considerable damage around the dents, which was more severe for the case using grease G1.

These results show that the EP additive in grease G3 ally to a higher viscosity of the base oil probably plays an important role in protecting the surface against fatigue damage. Results obtained using greases G1 and G2 (with a similar base oil viscosity) reveal that the difference in soap concentration might be the key to justify the different results obtained. Apparently, an increasing in soap concentration contributes to a better protection against damaged by contact fatigue.

Conclusions

The use of artificial dents proved to be useful in localizing and accelerating damage by rolling contact fatigue. It also allowed the comparison of the damaged area growth and the wear volume evolution for the different lubricants.

Concerning the influence of grease composition on the damage by fatigue it can be concluded that:

- An increase in oil viscosity results in a higher value of λ and consequently in a smaller damaged area around the dents and less wear. This was proven by the tests using the base oils.
- An increase in % of soap concentration benefits the protection of the surfaces against contact fatigue damage. This was proven by the tests using greases G1 (9 % soap concentration) and G2 (12 % soap concentration) that exhibit a similar base oil viscosity but different % of soap concentration. The values of the damaged area around the dents were higher for the tests using G1.
- Finally, the existence of an additive, such as EP, in grease G3, ally to a higher viscosity of the base oil, lead to an almost inexistent damaged in the end of the tests. In this case it is believed that the influence of the additive was more important than the influence of the other two parameters (base oil viscosity and % of soap concentration).

These conclusions on the influence of grease composition are consistent with the results that were obtained by several authors that studied the influence of grease composition on the film thickness (see Chapter II, Section 4.5.2.). It is believed that under fully-flooded conditions (those that existed in the experimental tests) an increase in base oil viscosity and % of soap composition leads to a greater film thickness and consequently to a better protection of the surfaces against contact fatigue.

Chapter V

Residual stresses evaluation

1. Introduction

This chapter is divided into two parts: the first one concerns residual stresses measurements performed in a disc, before and after the experimental tests; and the second one presents a numerical study, using the FEM, to evaluate the residual stress field due to an artificial dent.

Residual stresses measurements were performed at DF of the FCTUC using the X-ray diffraction technique. The goal is to have an idea of how the residual stresses evolutes from the beginning to the end of the experimental tests.

The numerical study using FEM was performed using the commercial program ABAQUS. This study provides information on the residual stresses distribution generated by a dent printed in the surface of the disc. The results obtained are to be used in the calculation of SIFs, described in Chapter VI.

2. Residual stresses

2.1. Definition

Residual stresses are those stresses that remain in a material or body after manufacture and processing in the absence of external forces or thermal gradients. They can be defined as either macro or micro stresses and both may be present in a component at any one time. *Macro residual stresses*, which are often referred to as Type I residual stresses, vary within the body of the component over a range much larger than the grain size. *Micro residual stresses*, which result from differences within the microstructure of a material, can be classified as Type II or III. Type II residual stresses are micro residual stresses that operate at the grain-size level; Type III are generated at the atomic level [82].

Micro residual stresses often result from the presence of different phases or constituents in a material. They can change sign and/or magnitude over distances comparable to the grain size of the material under analysis.

The different types of residual stress are shown schematically in Figure 5.1.

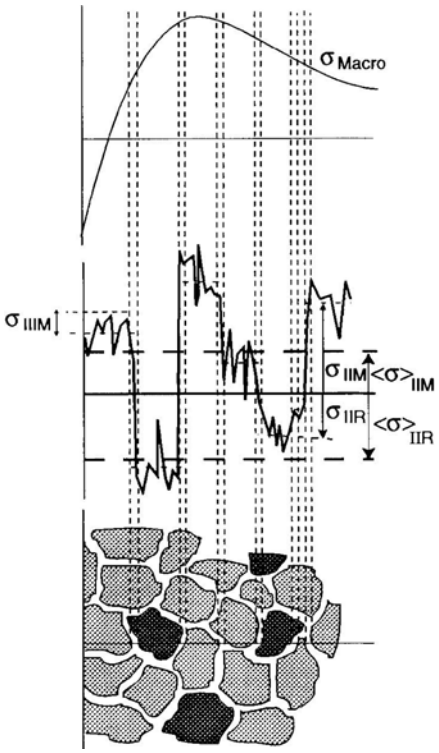


Figure 5.1 – Categoritation of residual stresses according to length scales [83].

2.2. Origins

Residual stresses develop during most manufacturing processes involving material deformation, heat treatment, machining or processing operations that transform the shape or change the properties of the material.

Both the magnitude and distribution of the residual stress can be critical to the performance of a component. Tensile residual stresses in the surface of a component are generally undesirable since they can contribute to, and are often the major cause of, fatigue failure, quench cracking and stress-corrosion cracking. Compressive stresses in the surface layers are usually beneficial since they increase both fatigue strength and resistance. In general, residual stresses are beneficial when they operate in the plane of the applied load and are opposite in sense.

The origins of residual stresses in a component can be classified as: 1) mechanical: as a result of manufacturing processes that produce non-uniform plastic deformation. They may develop naturally during processing or treatment, or may be introduced deliberately to develop a particular stress profile in a component; 2) thermal: due heat treatments or to non-uniform heating or cooling operations; and 3) chemical: due to chemical changes in the material.

3. Residual stress analysis

3.1. X-ray diffraction technique

To perform the analysis of the residual stress field in the discs, due to heat treatment, the *x-ray diffraction technique* was used. This technique relies on the elastic deformation within a polycrystalline material to measure internal stresses in a material. The deformations cause changes in the spacing of the lattice planes from their stress free value to a new value that corresponds to the magnitude of the applied stress. This new spacing will be the same in any similarly oriented planes, with respect to the applied stress and the crystal lattice therefore effectively acts as a very small strain gauge.

The measurement itself is relatively straightforward and equipment readily available. During measurement the specimen is irradiated with high energy X-rays that penetrate the surface, the crystal planes diffract some of these X-Rays, and a detector, which moves around the specimen to detect the angular positions where diffracted X-rays are located, records the intensity of these rays at that angular position. The location of the peaks enables the user to evaluate the stress within the component.

For each inclination of the sample, defined by two angles ϕ and ψ, the strain ε_n in the direction normal to the diffracting planes, is thus related to the corresponding lattice spacing d_n (see Figure 5.2). The distance d_n is also linked directly to the diffraction angle θ_n by Bragg's law, as it is expressed in equation.

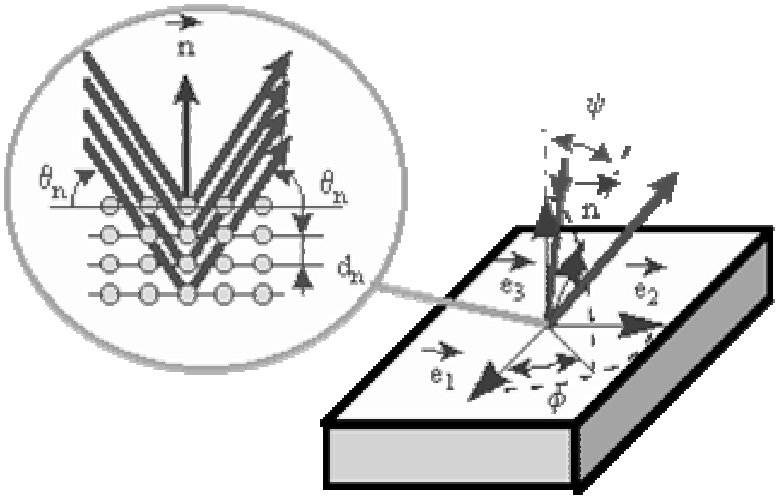


Figure 5.2 – Residual stresses measurement using the x-ray diffraction technique [84].

$$\left.\begin{array}{l} \varepsilon_n = \dfrac{d_n - d_0}{d_0} \\ \\ \lambda = 2d_n \sin\theta_n \\ \lambda = 2d_0 \sin\theta_0 \end{array}\right\} \quad \varepsilon_n = \frac{\sin\theta_0 - \sin\theta_n}{\sin\theta_n} \qquad (5.1)$$

Among the several experimental methods that can be used to evaluate the stresses within a material using this diffractometer technique, the $Sin^2\ \psi$ method was used. This method has the advantage that inclined measurements are made at a number of angles ψ rather than at only one.

A relation between the strain ε_n and the stress tensor components is derived through a micro mechanical approach of the material. Using the equilibrium conditions of the free surface (plane stress), the following equation, called $Sin^2\ \psi$ law, is obtained:

$$\varepsilon_n = \frac{1}{2} S_{2(hkl)} \sigma_\phi Sin^2\psi + \varepsilon_1 \qquad (5.2)$$

The plot of $\varepsilon_n = f(sin^2\psi)$ gives a straight line (Figure 5.3), whose slope allows computing the stress component σ_ϕ, in the direction ϕ of the surface, if the elastic constant $½S_{2(hkl)}$ is well known.

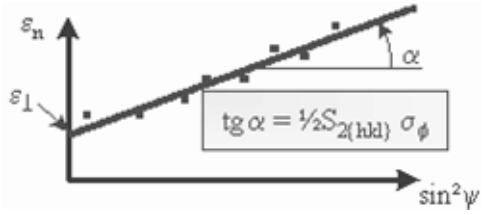


Figure 5.3 – Plot of $\varepsilon_n = f(sin^2\psi)$ [84].

3.2. Residual stresses measurements

Measurements on the discs were made on the longitudinal (σ_{11}) and transversal (σ_{22}) directions as shown in Figure 5.4.

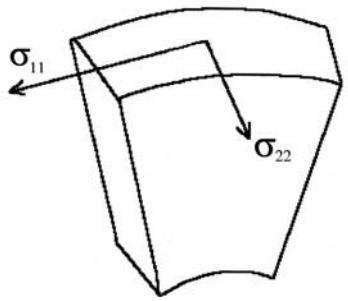


Figure 5.4 – Directions of the measurements performed.

All the measurements were performed in an intermediate position between the artificial dents, to avoid an influence of the residual stresses field generated by these dents.

The irradiated surface with high energy X-Ray was about 1.5×1.5 mm, which means that in this area an average value of the residual stresses is obtained.

Measurements on a disc before the tests

Measurements were performed inside the contact zone. The normal stress measured in the longitudinal and transversal directions of the disc were respectively $\sigma_{11} = -420$ MPa and $\sigma_{22} = -660$ MPa.

Measurements on a disc after the tests

Measurements were performed only on the longitudinal (σ_{11}) direction of the disc for different positions relatively to the contact center (Figure 5.5).

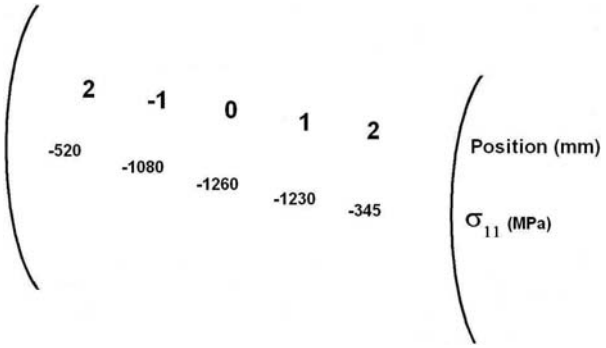


Figure 5.5 – Results of the longitudinal residual stresses (σ_{11}) for each measured zone.

The results and descriptions of the measured zones for each position are presented in Table 5.1.

Table 5.1 – Results of the longitudinal residual stresses descriptions of the measured zones for each position.

Position (mm)	Description of the measured zone	Residual stresses (MPa)
– 2	Outside of the contact zone	– 520
– 1	Half part inside the contact zone and the other half outside of it.	–1080
0	Inside the contact zone	– 1260
1	Half part inside the contact zone and the other half outside of it.	– 1230
2	Outside of the contact zone	– 345

Figure 5.6 presents the distribution of the longitudinal residual stresses. As it can be observed, a maximum is reached in the zone on the center of the contact, where the maximum Hertzian pressure moves.

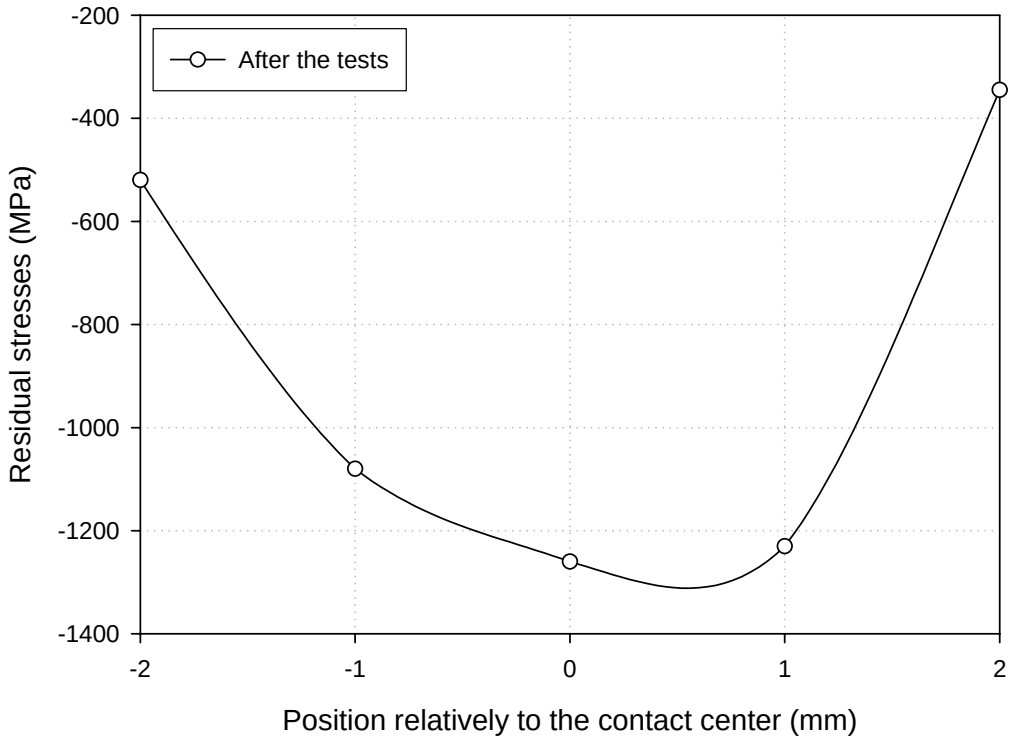


Figure 5.6 – Residual stresses (σ_{11}) distribution along the measured positions.

4. Residual stress analysis of an artificial dent using the FEM

Local residual stresses of the material in the dent's vicinity are not readily available from measurement using the X-ray technique. Therefore it was necessary to simulate the indent process in order to estimate the residual stresses.

4.1. Tensile test

To perform a finite element analysis it is necessary to determine the mechanical properties of the material (in this case the AISI 52100 steel) in its final stage (yield stress s_y, rupture stress, s_r and Young modulus E). Therefore, two tensile tests (T_1 and T_2) were carried out at LET using a test specimen with the geometry presented in Figure 5.7.

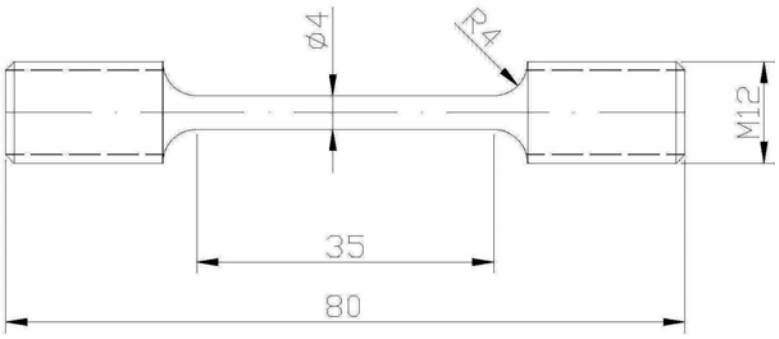


Figure 5.7 – Tensile test specimen geometry.

Both specimens fractured in the same manner, at a 45º shear plane. As example specimen T_1, after the tensile test, is presented in Figure 5.8.

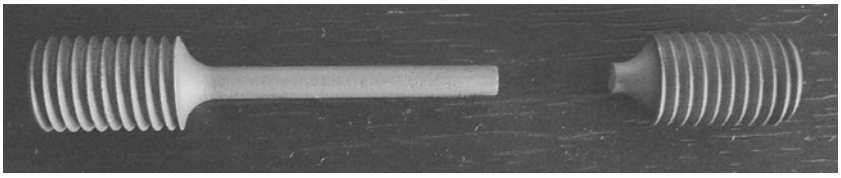


Figure 5.8 – Tensile specimen T_1 after test.

The tensile tests results average value m_s and standard deviation s_s, are presented in Table 5.2.

Table 5.2 – Material properties obtained from the tensile tests.

Specimen	s_y [MPa]	s_r [MPa]	E [GPa]
T_1	1700.0	2022.6	211.7
T_2	1794.0	2189.9	203.1
m_s	1747,0	2106,2	207,4
s_s	66,5	118,3	6,1

The records of σ = f (ε) obtained for the tensile test of specimens T_1 and T_2 are presented in Figures 5.9 and 5.10, respectively. The line in the figures represents the 0.2% proof stress that is the stress at which the material undergoes a 0.2% non-proportional (permanent) extension during the tensile test.

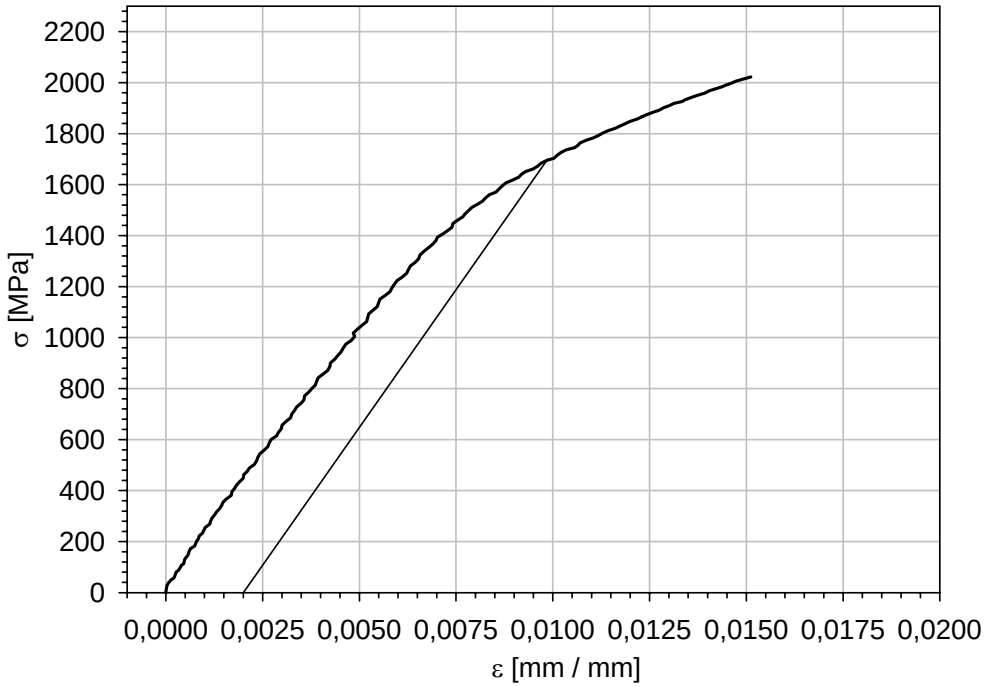


Figure 5.9 – Tensile test for the specimen T_1.

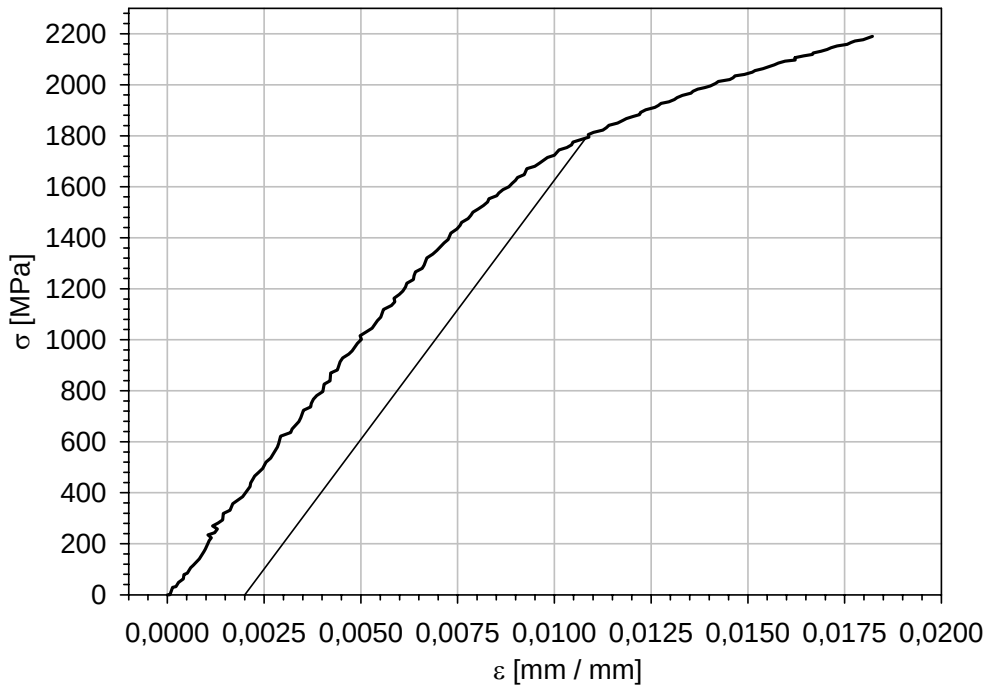


Figure 5.10 – Tensile test for the specimen T_2.

The results obtained from the tensile tests reveal that the AISI 52100 steel used has a brittle behaviour and a relatively high yield and rupture stresses.

4.2. Finite element analysis

A finite element model of a rigid indenter with a cone angle of 120º indenting a homogeneous half-space was used to determine the residual stress field. The finite element (FE) simulations were performed using the ABAQUS commercial FE code.

4.2.1. Finite element model

The problem is axisymmetric since the indenter and the facing body has a common axis of revolution. The FE mesh used for indent simulation is presented in Figure 5.11. In the contact zone the mesh is more refined and has a size that is approximately 0.019 μm, which is small when compared to the dent diameter (468 μm for a load of 2500 N). The element type used was a 3-dimensional continuum element with 8 nodes (see Figure 5.12). The facing body was divided into four parts due to its axis-symmetry and was meshed with 6071 elements and 10043 nodes.

The indenter was not a perfect cone but had a slight curvature at the end, because only with this change it was possible running the simulation. Obviously some errors can derive from this change, especially taking into account the fact that the indenter used on the experimental indents had not this curvature.

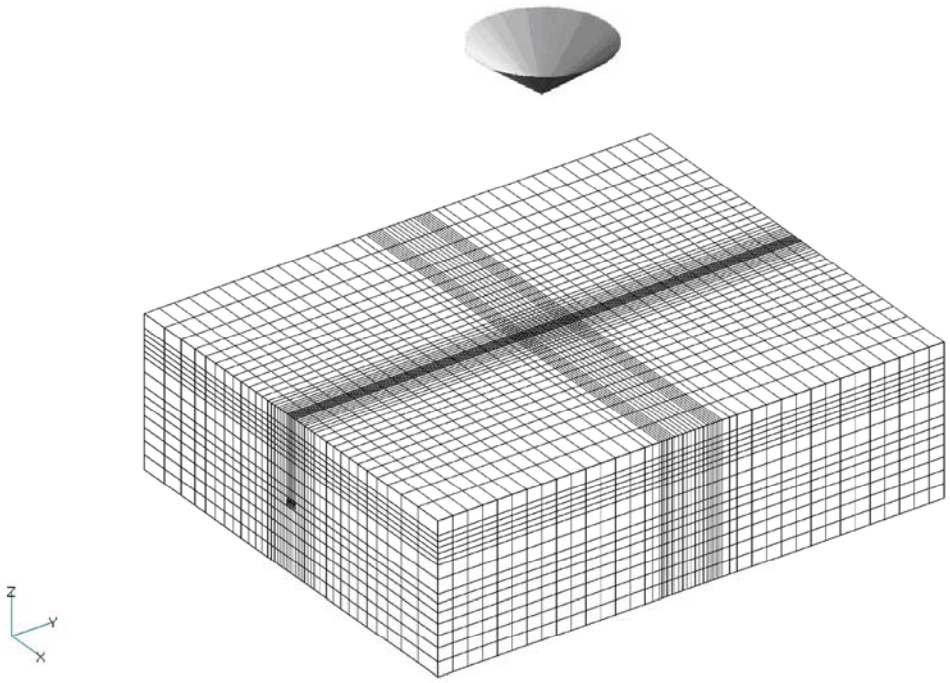


Figure 5.11 – Finite element mesh used for the indent simulation.

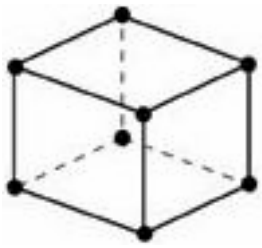


Figure 5.12 – Linear element (8-node brick) used in the FE analysis.

Boundary conditions were applied through half-infinite elements in order to simulate the half-space body (the nodes in the boundary are unable to move) and the load was applied to the indenter that penetrates the body.

To attest the quality of this numerical simulation, the real geometry of the dent obtained by experimental measurements (see Chapter III, Section 4) was

compared with the one obtained from the simulation. This comparison was made for the indenter penetrating the surface for different normal loads (1500, 2000 and 2500 N). The results are presented in Figures 5.13-5.15.

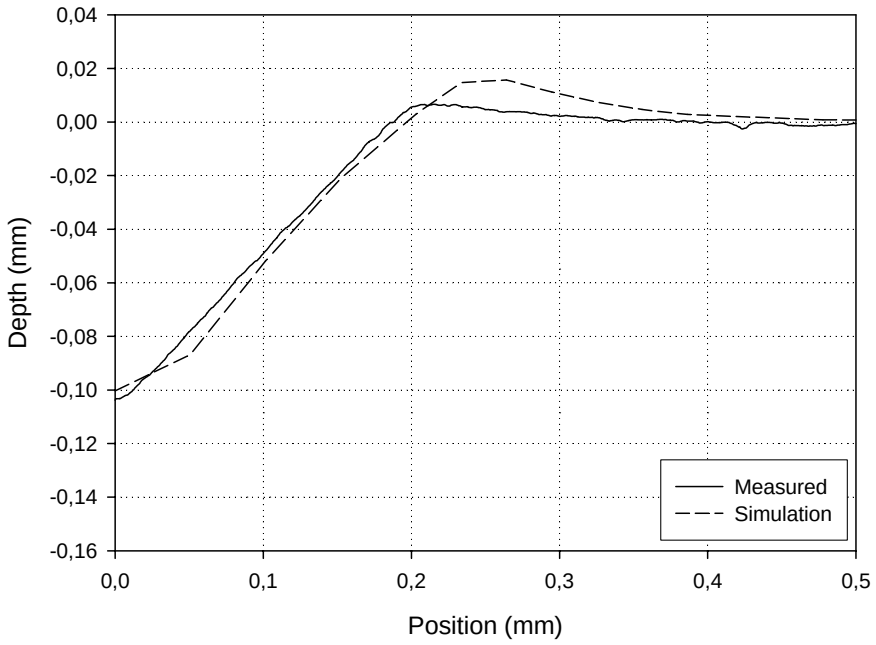


Figure 5.13 – Comparison between measured and simulated dent profiles for a 1500 N load.

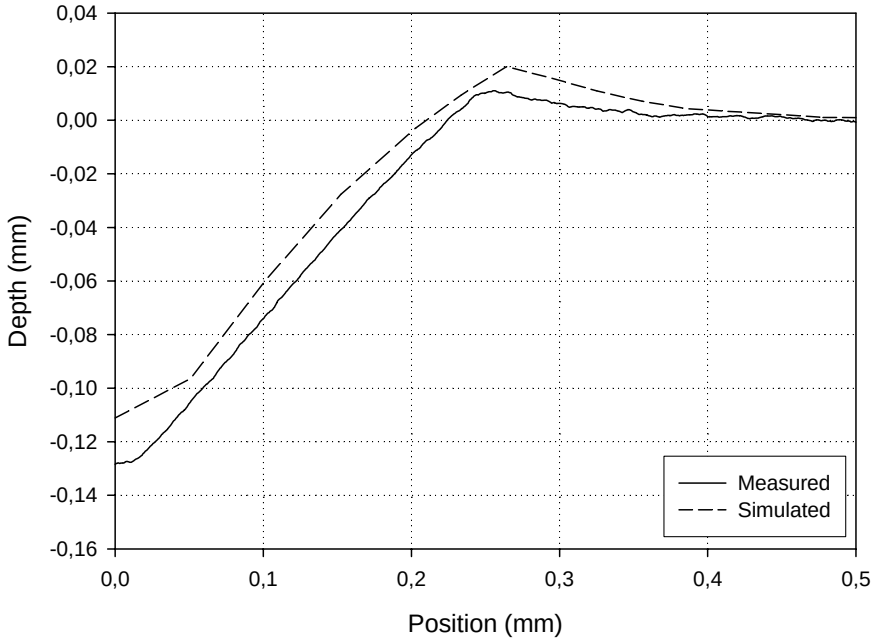


Figure 5.14 – Comparison between measured and simulated dent profiles for a 2000 N load

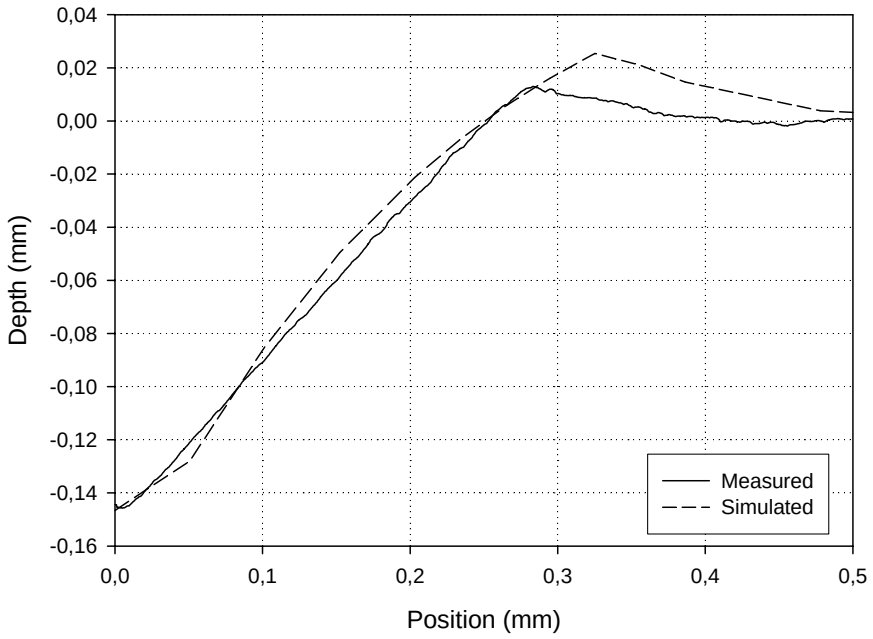


Figure 5.15 – Comparison between measured and simulated dent profiles for a 2500 N load.

It can be noted that the FE simulation gives a dent profile that is globally close to the measured one although a lack of good agreement can be observed, particularly around the dent shoulder. This discrepancy can be justified based on the errors that affect the experimental measurement and also due to the fact that the simulated indenter is not a perfect cone but has a slight curvature at the end, as it was referred above.

4.2.2. Results of the finite element analysis

Next, the results of residual stress and strain for a 2500 N normal load applied, equal to the one used to perform the dents for the experimental tests, are presented.

Figure 10 shows the three-dimensional residual stress field in terms of Von Mises stresses for the entire body after unloading. In Figure 5.16, a detailed image focusing the indent area can be seen.

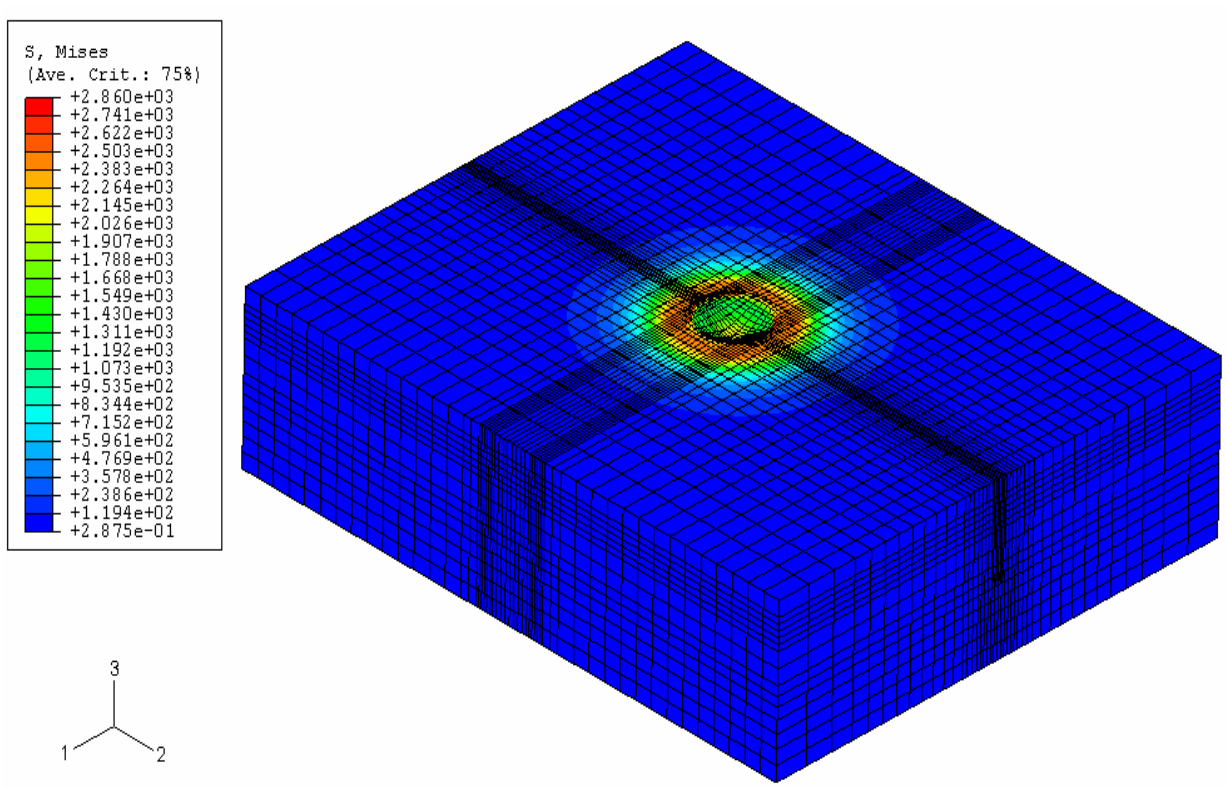


Figure 5.16 – Von Mises stresses after unloading (three-dimensional view).

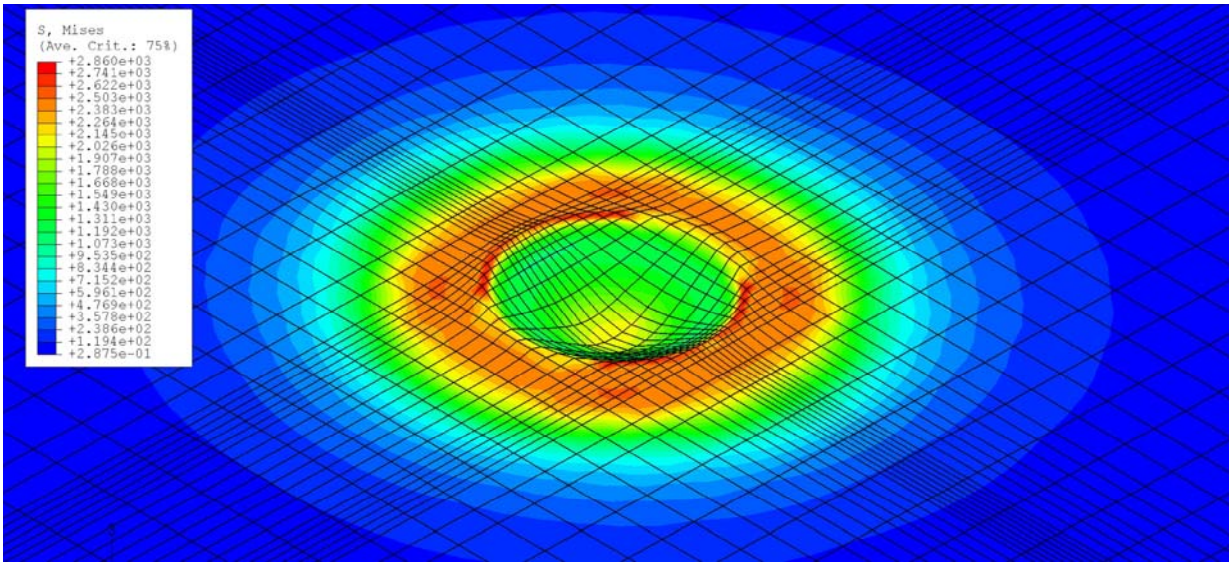


Figure 5.17 –Von Mises stresses after unloading (detailed three-dimensional view).

The equivalent plastic strain introduced during the indent process is given in Figure 5.18. The depth corresponds to the distance from the deformed surface. The horizontal axis corresponds to the distance from the center of the dent.

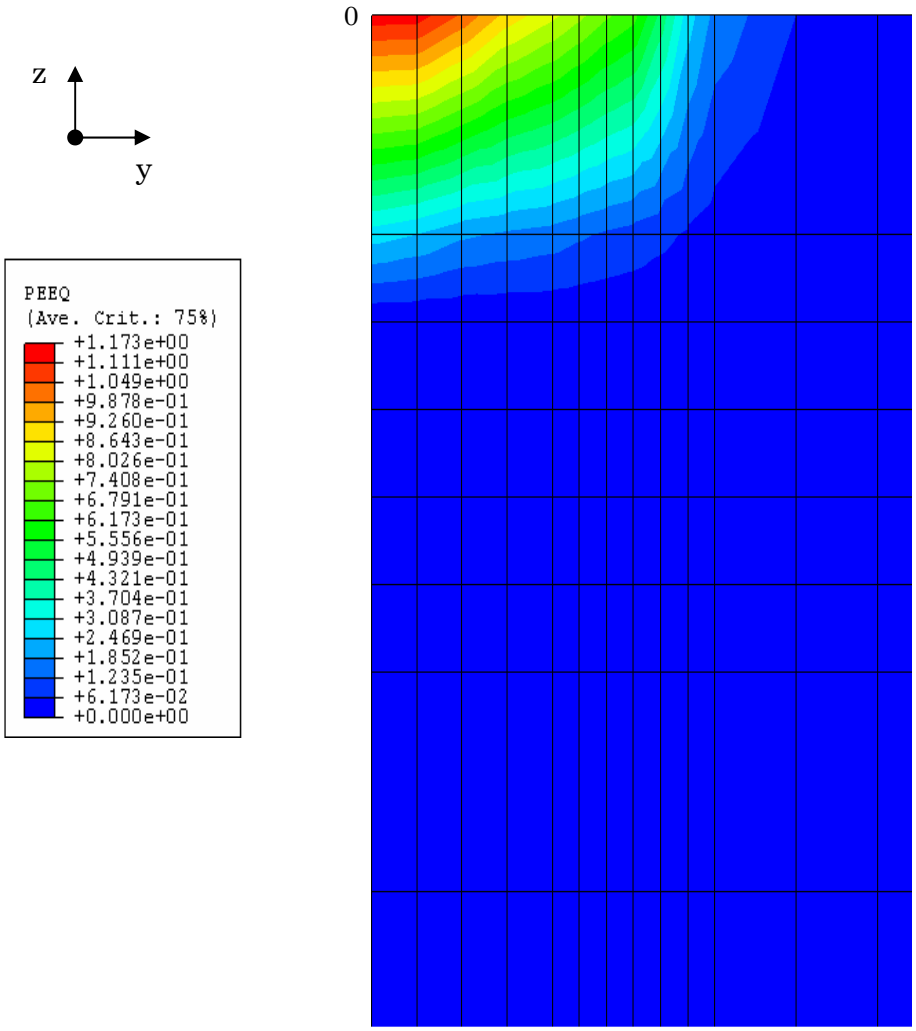


Figure 5.18 – Equivalent plastic strain introduced during the indent process.

Figure 5.19 shows the residual stress field after unloading. At the dent's shoulder very high Von Mises stresses are observed. These are always above 2000 MPa and up to 2860 MPa. Furthermore the stressed zone is not limited to the plastically deformed zone, but extends far away from it.

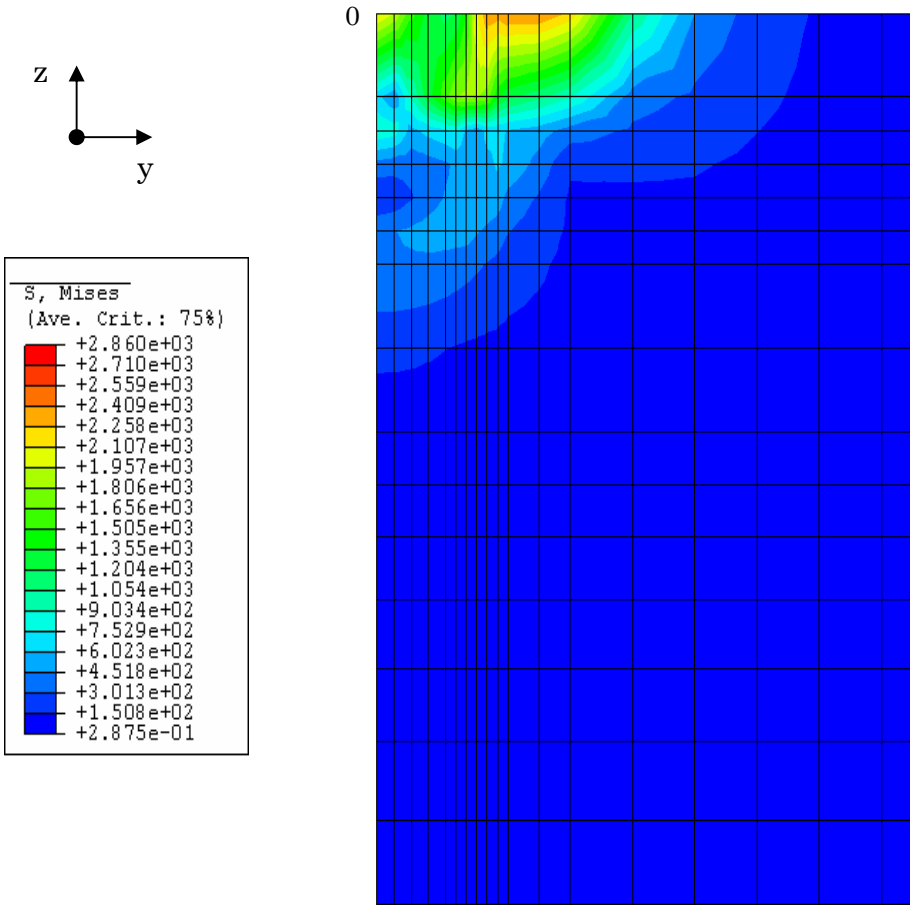


Figure 5.19 – Von Mises residual stress introduced during the indent process.

To better understand the role of residual stresses produced by the indent process, the principal stresses σ_{11}, σ_{22} and σ_{33} were considered as illustrated in Figures 5.20 and 5.21. Due to the symmetry, the principal stresses σ_{22} and σ_{33} are the same. Observing these figures it can be concluded that the stresses in the dent shoulder's zone are essentially compressive.

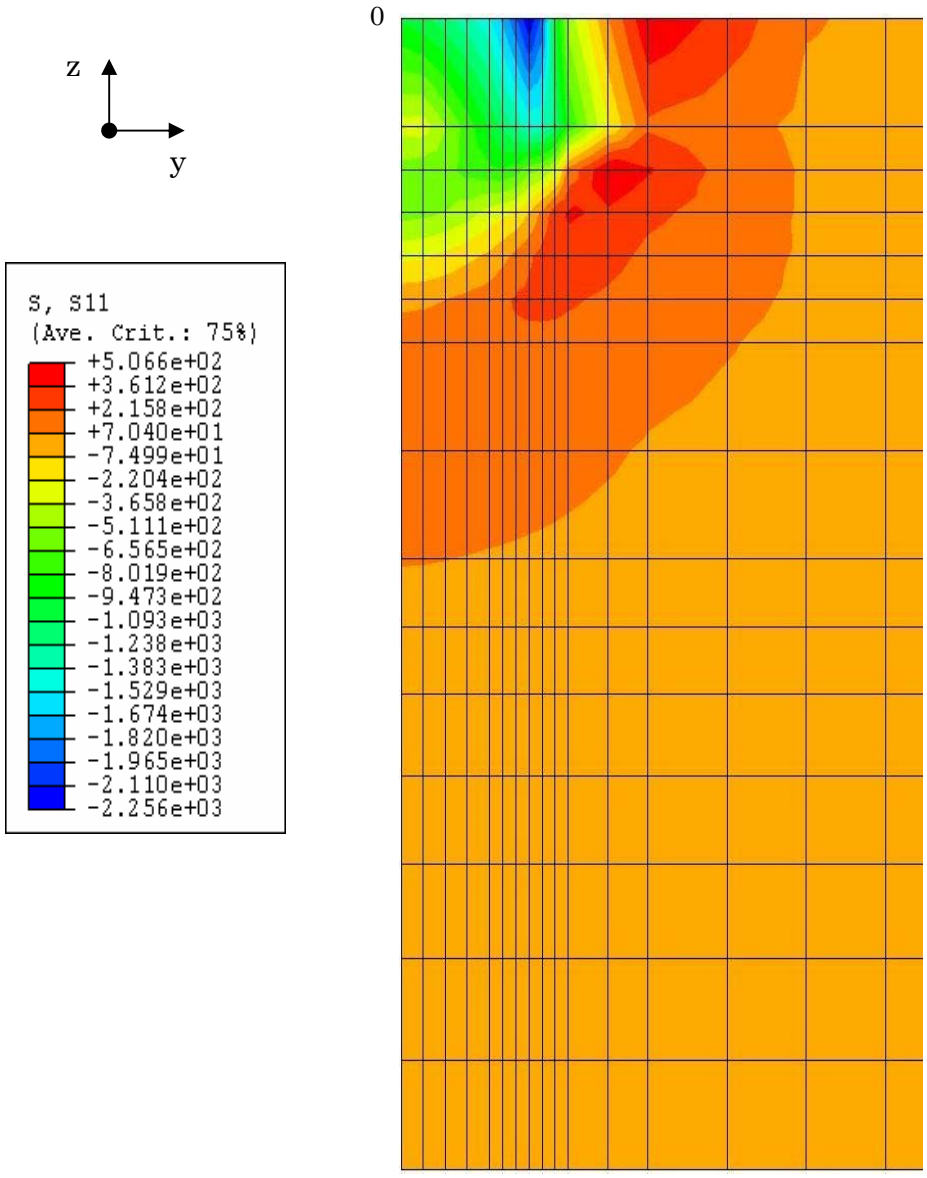


Figure 5.20 – Component σ_{11} of the residual stress tensor.

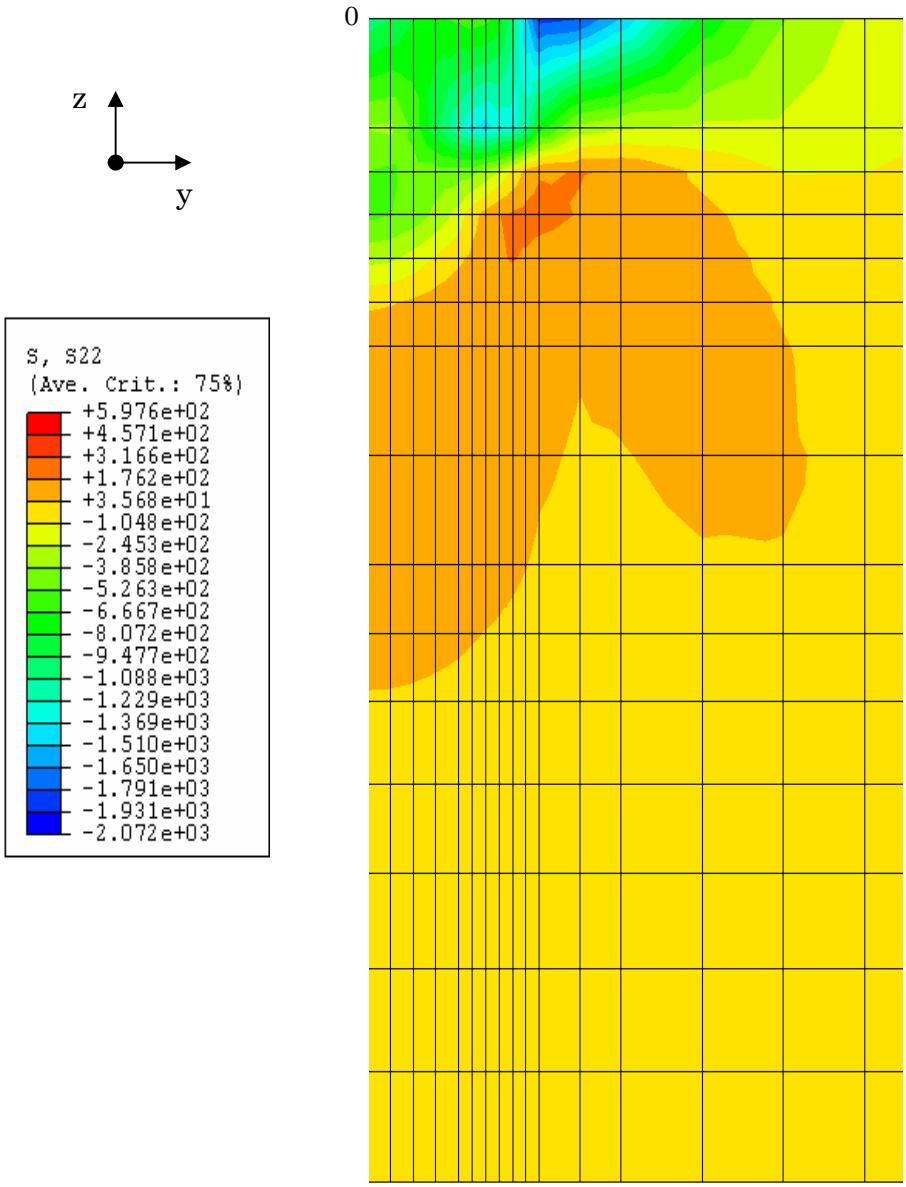


Figure 5.21 – Component σ_{22} and σ_{33} of the residual stress tensor.

Conclusions

Measurements using the X-ray technique on a disc before the experimental tests showed that the heat treatment introduced a residual stress in the centre of contact of σ_{11} = – 420 MPa and σ_{22} = – 660 MPa measured respectively on the longitudinal and transversal directions of the disc. Measurements performed after the tests revealed that for the center of the contact the longitudinal residual stress (σ_{11}) was considerably higher and equal to – 1260 MPa. Although these results are the results of an average value in the irradiated area, it can be concluded that in the end of the tests, a considerable increase on the residual stresses exists. This was probably due to high plastic deformation that occurred in the contacting zone, particularly in the center of the contact, where the Hertzian pressure distribution reaches its maximum value. Actually, measurements performed on the longitudinal (σ_{11}) direction of the disc for different positions relatively to the contact center showed that the residual stress value decreased as the position was being moved away from the center of the contact.

A numerical simulation of a dent similar to the one used for the experimental tests was carried out. Comparison between experimental measurements and numerical results showed that although small discrepancies exist a good agreement can be found.

The results for equivalent plastic strain and residual stresses field after unloading showed that high values of both plastic strain and residual stresses were introduced. It was found that the higher values were situated at the dent's shoulder, which confirms the experimental results presented in Chapter IV – it was precisely at the dent's shoulder that the initial damage by fatigue took place.

The results from this simulation, particularly the residual stress distribution at the surface corresponding to the principal stress σ_{33} (stress field in the direction of the applied load) is used in the Chapter VI to assess the influence of residual stresses due to a dent in the SIF distribution for a surface-breaking inclined crack that initiates at the dent's shoulder (where the residual stress reaches its maximum value).

Chapter VI

Stress intensity factors calculation

1. Introduction

In this chapter, stress intensity factors (SIFs) calculation at the tip of an inclined fluid-filled surface-breaking crack initiated in an artificial dent under RCF conditions is presented.

SIF is a measure of the stress field at a crack tip. Its calculation is of major importance because it controls fatigue crack growth. To determine it several methods have been developed. Some of the most popular methods are listed below:

- Finite element analysis;
- Boundary element analysis;
- Green's function methods;
- Dislocation density and other related strain nuclei techniques.

The choice of a method depends on the application. In the case of inclined edge cracks, it was shown that a very efficient way of determining SIFs for arbitrary loadings is to use a Green's function technique [85]. Therefore this method was chosen in this work.

Based on the edge Green's function presented graphically by Rooke *et al.* [85], Fletcher and Beynon [38,86] presented a method of SIFs calculation method for inclined fluid-filled surface-breaking cracks under contact loading. Although this method is rather simple the results, when compared with those from more complex ones, show very good agreement [86].

This method considers the influence of the Hertzian pressure moving along the contact and the lubricant pressure inside the crack. However, to evaluate more accurately SIFs at the tip of crack initiated in an artificial dent, the residual stresses field evaluated by FEM analysis in Chapter V was also included in the method.

2. Stress intensity factor calculation model

Experimental investigations on RCF of artificial dented surfaces (see Chapter IV) showed that, under pure rolling conditions, crack initiation and propagation occurs at the base of the dent's shoulder when the shoulder is compressed by the Hertzian load after a given number of cycles (Figure 6.1).

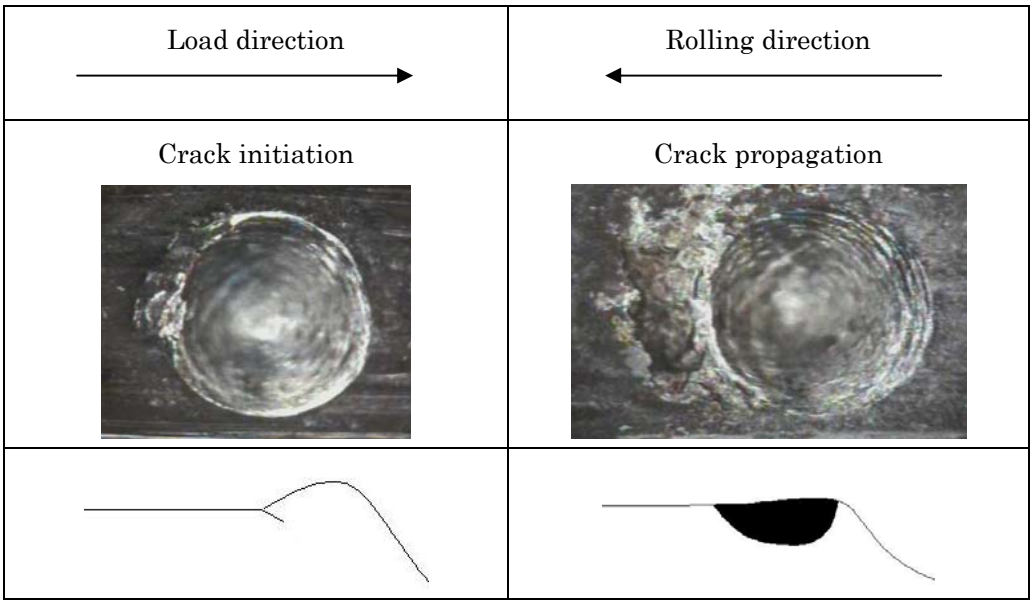


Figure 6.1 – Crack initiation and propagation in a dent.

Let us consider an infinitely wide inclined edge crack in an elastic half space, initiated at the base of an dent's shoulder and loaded by a compressive Hertzian normal pressure distribution (Figures 6.2 and 6.3). The residual stress distribution that results from the artificial dent is considered and the fluid pressure at the crack mouth is assumed to be equal to the Hertzian contact pressure at the crack mouth. When the contact does not cover the crack the fluid pressure is assumed to be equal to zero.

The crack is assumed to remain open throughout the passage of the load from the left to the right of the half space. Movement in this direction is consistent with the assumption that the crack will remain open and is filled with fluid at all times, and also represents the condition in which rolling contact fatigue crack development is found to occur [87]. Any combination of surface load, crack length and contact half-width for which these assumptions are appropriate may be considered.

The method makes use of separate solutions available for the SIFs due to the edge load, the residual stress field and the pressurized fluid in the crack, and superposition of these solutions is used to describe the combined system.

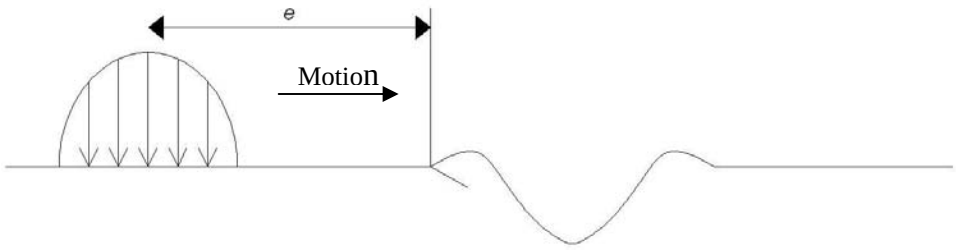


Figure 6.2 – An inclined edge crack in a half-space initiated at the base of an artificial dent's shoulder and loaded by a Hertzian contact moving from left to right.

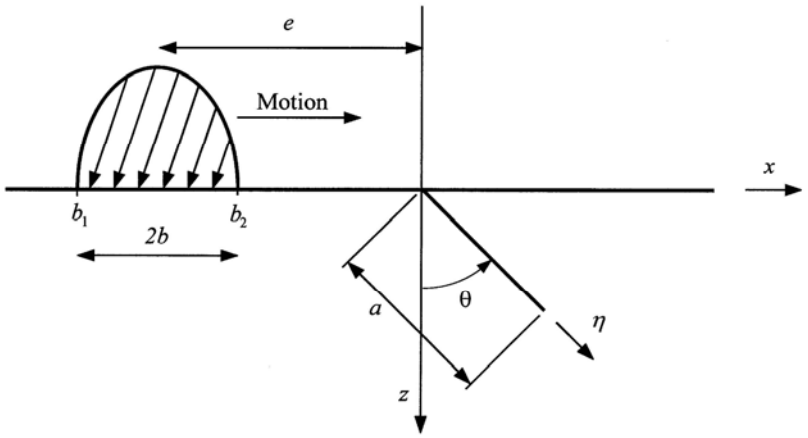


Figure 6.3 – Sheme of an inclined edge crack in a half-space [38].

2.1 Modeling the moving Hertzian load

The Hertzian contact half-width b and the position of its centre e are used to define the boundaries of the contact b_1 and b_2:

$$b_1 = e - b$$
$$b_2 = e + b \tag{6.1}$$

Movements of the centre of the contact in the range $e = \pm 3b$ is accomplished by variation in e in increments of $b/10$. The normal contact stress has a Hertzian distribution [6] and is defined for $b_1 \leq x \leq b_2$ as:

$$P(x) = -P_0 \left\{ 1 - \left[\frac{2(x-e)}{b_2 - b_1} \right]^2 \right\}^{1/2} \tag{6.2}$$

where P_0 is the maximum Hertzian contact pressure. The tangential contact stress depends on the friction coefficient between the surfaces μ:

$$Q(x) = -\mu P(x) \tag{6.3}$$

In this work the tangential contact stress will not be considered due to the fact that all the experimental tests were carried out under pure rolling contact conditions.

SIFs due to the Hertzian load are calculated using the edge Green's functions presented graphically by Rooke *et al.* [85]. These functions are shown in Figures 6.4 and 6.5.

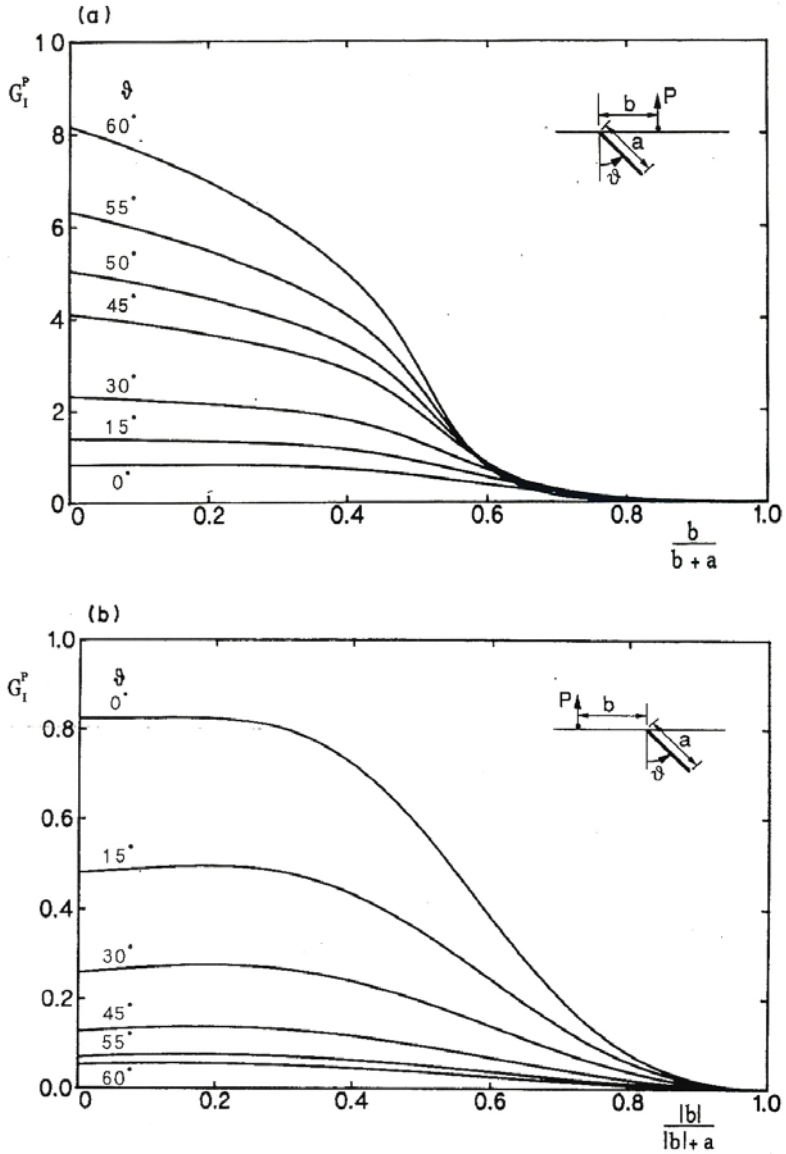


Figure 6.4 – Mode I edge Green's functions due to a point force normal to the edge of the half-plane for (a) $b > 0$ and (b) $b < 0$ [85].

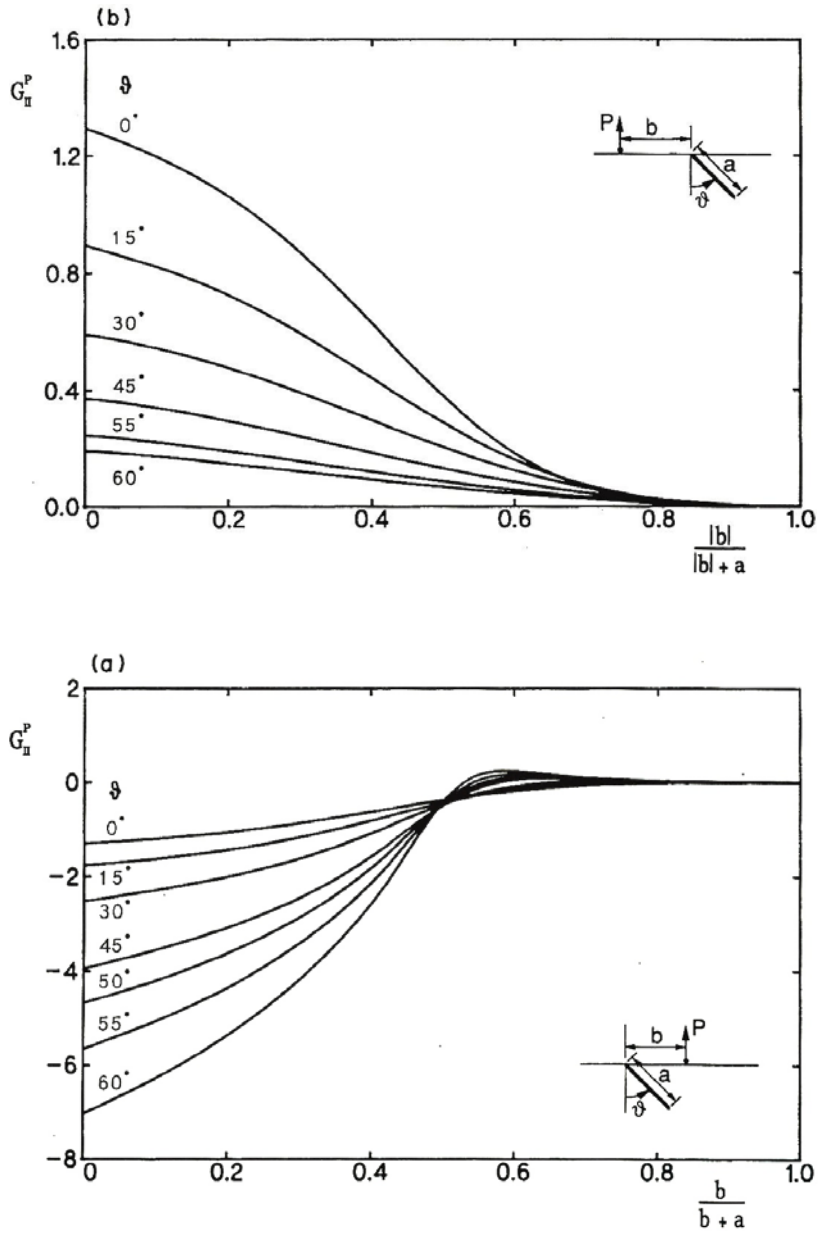


Figure 6.5 – Mode II edge Green's functions due to a point force normal to the edge of the half-plane for (a) $b > 0$ and (b) $b < 0$ [85].

For each crack growth angle, 4 functions are required to describe the influence of normal load on the opening mode (K_I) and shearing mode (K_{II}) SIFs. The curves were described by the function

$$G_N^{PH}(Y) = \left(1 - \left(\frac{x}{x+a}\right)^{u_N^{PH}}\right)^{v_N^{PH}} \sum_{i=0}^{5} n_{Ni}^{PH} \left(\frac{x}{x+a}\right)^i \tag{6.4}$$

where N = *(I, II)*, H = (left, right), a is the crack length and x is a general position on the surface. The constants u_N^{PH}, v_N^{PH} and n_{Ni}^{PH} were determined for each curve by a least-squares fit (see results in Table 6.1).

SIFs due to the residual stress distribution are found by numerical integration of the product of the appropriate Green's function and residual stress distribution, giving:

$$K_{N\,\text{Hertzian load}}^{PH} = \frac{1}{\sqrt{\pi a}}\left[\int_{l_2}^{l_2} P(x)G_N^{Pleft}(x)dx + \int_{l_3}^{l_4} P(x)G_N^{\text{Pr}\,ight}(x)dx\right] \tag{6.5}$$

The limits of integration are chosen for each contact position e and are defined as:

$$\begin{aligned} l_1 &= \min(0, b_2) \\ l_2 &= \min(0, b_1) \\ l_3 &= \max(0, b_2) \\ l_4 &= \max(0, b_1) \end{aligned} \tag{6.6}$$

Table 6.1 – Edge Green's function parameters for cracks at different angles [38].

θ	Green's function	u_N^{PH}	v_N^{PH}	n_{N0}^{PH}	n_{N1}^{PH}	n_{N2}^{PH}	n_{N3}^{PH}	n_{N4}^{PH}	n_{N5}^{PH}
30º	G_I^{left}	7,160	9,467	2,415	-2,303	10,37	-18,46	-17,25	31,05
	G_I^{right}	0,850	1,932	0,268	1,074	-0,309	4,707	-7,365	2,844
	G_{II}^{left}	6,660	47,06	-2,549	2,587	-3,732	38,03	-96,28	88,90
	G_{II}^{right}	1,329	4,588	0,583	0,418	4,489	-2,075	-1,077	16,44
45º	G_I^{left}	2,557	8,452	4,140	-0,265	-14,23	101,9	-56,52	1,208
	G_I^{right}	0,233	1,064	0,206	1,900	-3,194	8,215	-13,57	7,020
	G_{II}^{left}	4,932	55,35	-3,926	1,646	18,06	-21,87	-102,2	139,9
	G_{II}^{right}	1,624	5,242	0,367	-0,280	3,933	-5,242	-1,998	18,45
60º	G_I^{left}	5,070	38,65	8,163	-4,452	-14,55	27,78	43,45	32,05
	G_I^{right}	2,391	-0,127	0,066	0,011	-0,179	0,186	-0,227	0,145
	G_{II}^{left}	7,333	460,7	-7,085	9,689	-31,63	178,9	-224,6	-203,7
	G_{II}^{right}	1,374	3,328	0,187	0,032	-0,076	3,037	-5,772	3,472

2.2 Modeling the fluid pressure inside the crack

In addition to edge Green's functions, Rooke et al. [85] presented crack line Green's functions for calculation of SIFs due to a distribution of stress along the crack line. In Figures 6.6 (a) and (b) the results for the crack-line Green's functions along the whole length of the crack are presented.

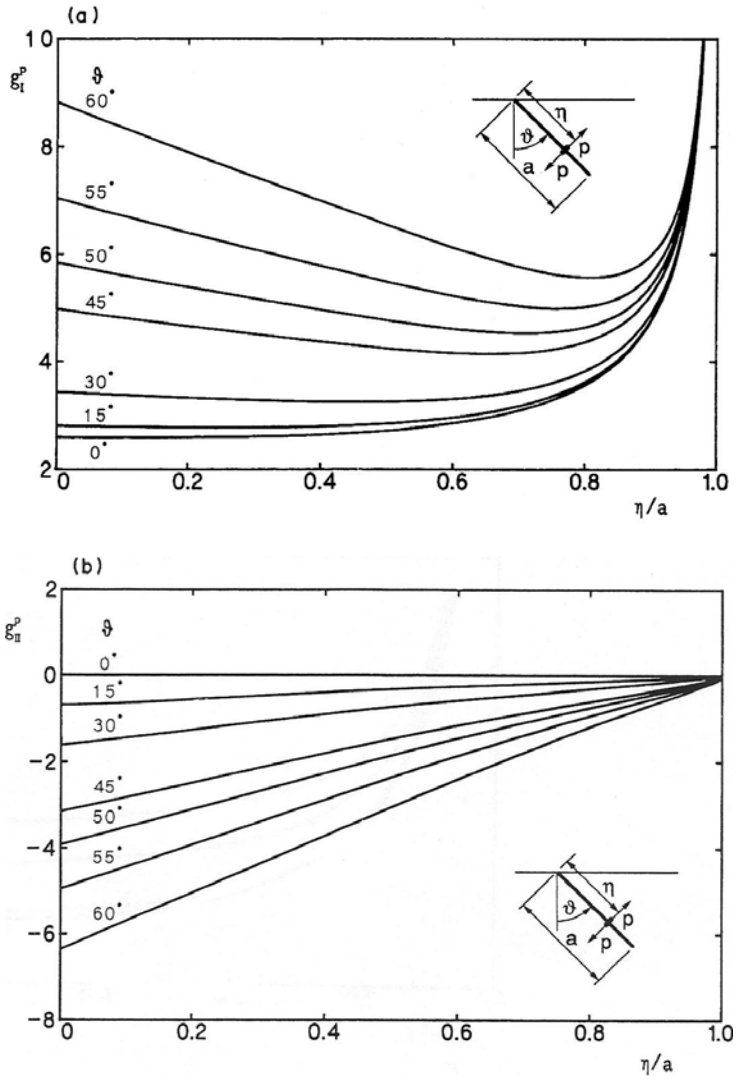


Figure 6.6 – Crack-line Green's functions due to a point force normal to the crack surfaces for (a) Mode I and (b) Mode II [85].

The process that conduces to the fluid trapped inside the crack is described in Figure 6.7. The fluid is considered to be incompressible and the pressure at the crack mouth is equal to the Hertzian pressure when it moves on the crack and zero when it does not moves on it. It is also considered that the fluid volume inside the crack does not change when the Hertzian pressure is moving on it.

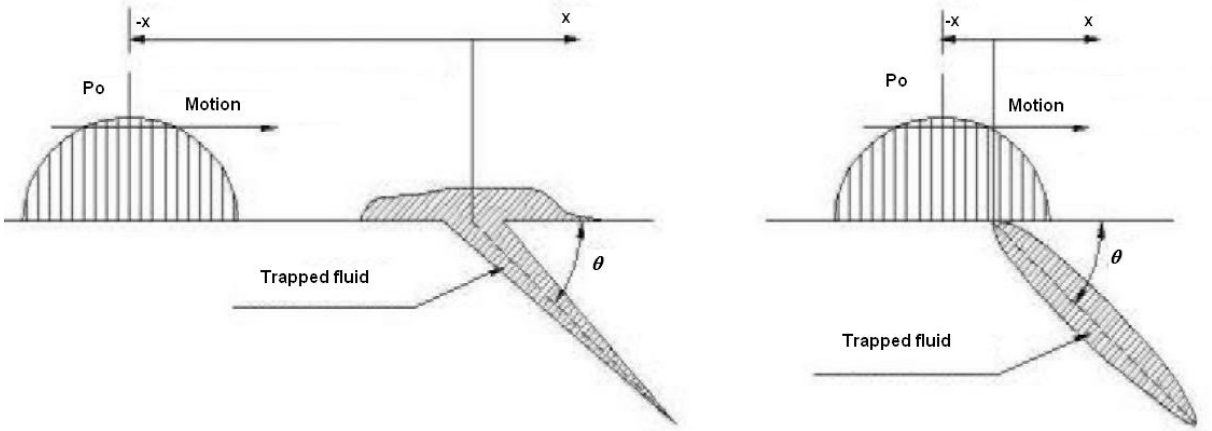


Figure 6.7 – Fluid trapped inside a crack (adapted from [88]).

Different distributions of fluid pressure inside the crack can be used. Benuzzi *et al.* [88] considered three distributions of fluid pressure inside the crack (see Figure 6.8) and concluded that the uniform distribution conduces to higher values of SIFs but it overestimates the influence of the fluid pressure inside the crack. On the other hand, the parabolic and linear distributions, like expectable, conduce to similar SIFs values and appears to be more realistic.

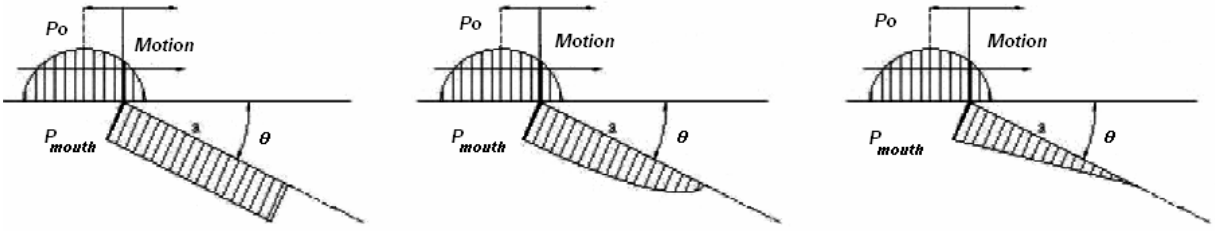


Figure 6.8 – Distributions of fluid pressure inside the crack: a) uniform, b) parabolic and c) linear (adapted from [88]).

In this work, it is assumed that the fluid inside the crack transmits stress normally to the crack faces, with a magnitude decreasing linearly from the crack mouth to its tip (linear distribution), defined by

$$P_{crack}(\eta) = P_{mouth}\left(1-\frac{\eta}{a}\right) \tag{6.7}$$

where P_{mouth} is the magnitude of the Hertzian contact pressure acting at the crack mouth, which is dependant on contact patch position e.

Rooke et al. [85] presented the crack line Green's functions in the form

$$g_I^P(X) = \frac{1}{\sqrt{1-X^2}} \sum_{i=0}^{5} p_i^I X^i \tag{6.8}$$

$$g_{II}^P(X) = \sqrt{1-X^2} \sum_{i=0}^{5} p_i^{II} X^i \tag{6.9}$$

where $X = \eta / a$ and p_i^I e p_i^{II} were presented as a table of constants (Table 6.2).

Table 6.2 – Coefficients of crack-line Green's functions [85].

θ	n	p_n^I	p_n^{II}
30º	0	3,4234	-1,6111
	1	-0,6050	1,7422
	2	-1,2806	-1,2033
	3	-0,7989	4,2372
	4	2,2170	-7,3834
	5	-0,9946	4,0384
45º	0	4,9822	-3,1416
	1	-1,7783	3,3760
	2	-1,2527	-2,5540
	3	-2,3988	6,7746
	4	3,6288	-9,4572
	5	-1,2132	4,4956
60º	0	8,8234	-6,3610
	1	-4,8736	7,0066
	2	-2,9592	-6,6356
	3	-0,5485	14,0194
	4	1,3147	-15,0334
	5	0,1735	6,1170

For each position of the Hertzian load $K_{N\ \mathrm{crack}}$ may be found by numerical integration using the same methods as $K_{Nsurface}^P$:

$$K_{Ncrack} = \frac{1}{\sqrt{\pi a}} \int_0^a P_{crack}(\eta) g_N^{Pcrack}\left(\frac{\eta}{a}\right) d\eta \tag{6.10}$$

where $N = (I, II)$.

2.3 Modeling the residual stress field due to an artificial dent

SIFs due to the residual stresses field are calculated using the edge Green's functions already presented. For each crack growth angle, 4 functions are also required to describe the influence of the residual stresses on the opening mode (K_I) and shearing mode (K_{II}) SIFs. The curves were described by the function

$$G_N^{RH}(Y) = \left(1 - \left(\frac{x}{x+a}\right)^{u_N^{PH}}\right)^{v_N^{PH}} \sum_{i=0}^{5} n_{Ni}^{PH} \left(\frac{x}{x+a}\right)^i \tag{6.11}$$

where $N = (I, II)$, H = (left, right), a is the crack length and x is a general position on the surface. The constants u_N^{PH}, v_N^{PH} and n_{Ni}^{PH} are the same that were determined above and presented in Table 6.1.

The function *R(x)* represents the residual stress distribution of the principal stress σ_{33} at the surface (see Chapter V, Part 4.2.2.). This distribution is presented in Figure 6.9. On the other hand, Figure 6.10 shows an adapted axis of the distribution were the maximum stress value (where the crack is expected to initiate) coincides with x = 0. This adaptation was performed because to be coherent with the coordinates system presented in Figure 6.2.

The function *R(x)* had to be divided into several parts so that better accuracy could be achieved when fitting the curve with polynomial equations. *R(x)* is then defined for both sides of the crack by:

Left side of the crack:

$$-3 \le x \le -0.8 \quad \Rightarrow R^{left}(x) = 0$$

$$-0.8 \le x \le -0.3 \quad \Rightarrow R^{left}(x) = 1.02\times10^{9} + 3.90\times10^{12}\,x + 4.99\times10^{15}\,x^2 + 2.14\times10^{18}\,x^3$$

$$-0.3 \le x \le 0 \quad \Rightarrow R^{left}(x) = \frac{-1.27\times10^{9} - 9.60\times10^{12}\,x}{1 - 19539.59x}$$

Right side of the crack:

$$0 \le x \le 0.3 \quad \Rightarrow R^{right}(x) = -1.27\times10^{9} + 1.23\times10^{13}\,x - 1.44\times10^{16}\,x^2$$

$$0.3 \le x \le 0.6 \quad \Rightarrow R^{right}(x) = 9.39\times10^{6} + 4.98\times10^{12}\,x - 1.44\times10^{16}\,x^2$$

$$
\begin{aligned}
0.6 \le x \le 0.9 \qquad & \Rightarrow R^{right}(x) = \frac{6.62\times10^{8} - 9.07\times10^{11}x}{1-1822.35x} \\
0.9 \le x \le 1.4 \qquad & \Rightarrow R^{right}(x) = -8.91\times10^{9} + 3.97\times10^{13}x - 6.00\times10^{16}x^{2} \\
& \qquad + 3.81\times10^{13}x^{3} - 8.75\times10^{21}x^{4} \\
1.4 \le x \le 3 \qquad & \Rightarrow R^{right}(x) = 0
\end{aligned}
\tag{6.12}
$$

SIFs due to the residual stress distribution are found by numerical integration of the product of the appropriate Green's function and residual stress distribution, giving:

$$
K^{PH}_{Nresidualstress} = \frac{1}{\sqrt{\pi a}}\left[\int_{R} R(x)G_{N}^{Pleft}(x)dx + \int_{R} R(x)G_{N}^{\Pr ight}(x)dx\right] \tag{6.13}
$$

The limits of integration are defined, for both sides of the crack, according to the part of *R(x)* that is being integrated.

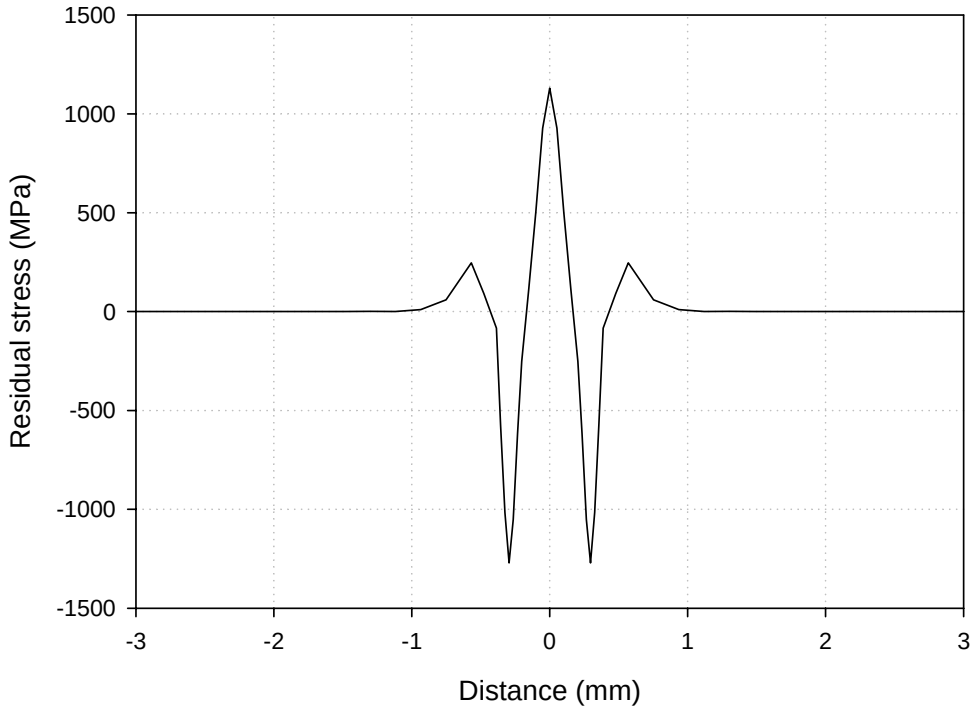


Figure 6.9 – Residual stress distribution due to an indent generated at the surface where x = 0 coincides with the dent's center.

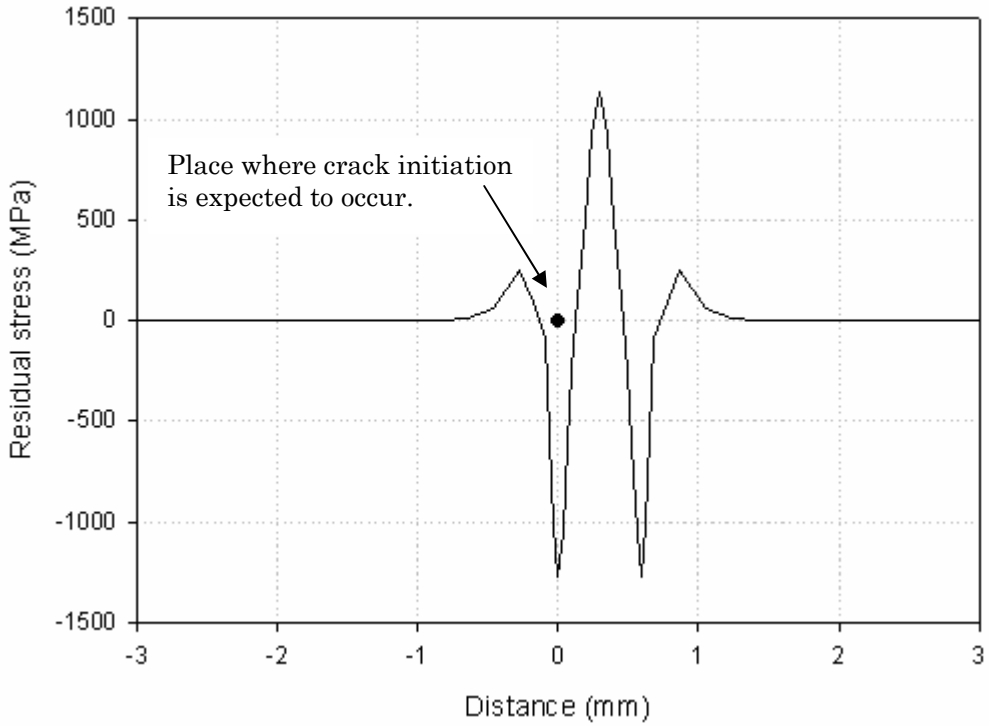


Figure 6.10 – Residual stress distribution due to an indent generated at the surface where x = 0 coincides with the maximum stress value.

3. Results

The resultant SIFs for the crack can then be found by superposition of the edge load, fluid pressure results and residual stresses, i.e.

$$K_N = K^P_{N\,\text{Hertzian load}} + K_{N\,\text{crack}} + K_{N\,\text{residual stresses}} \tag{6.14}$$

where $N = (I, II)$.

The SIFs K_I and K_{II} were calculated for cracks at angles 30, 45 and 60º to the surface normal, for a maximum Hertzian pressure P_0 = 2 GPa (the same used in the experimental tests) and for a crack length a_{crack} = 0.2 mm. The cracks growth angles were chosen based on the experimental results that showed that cracks tended to growth with an inclination that varied approximately between 30 to 60º. A very small crack length was chosen to simulate the crack initiation and so that the residual stress field can still be valid. A considerably bigger crack length initiated at the surface would induce changes in the residual stress field and consequently invalidating the analysis using the calculated residual stress field by FEM simulation.

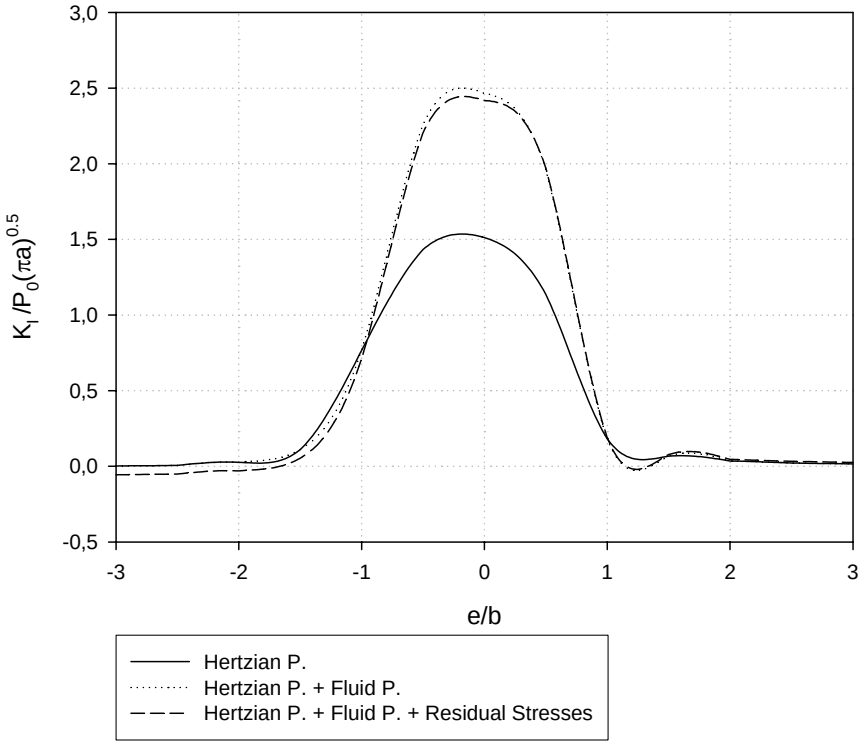
For each crack angle, SIFs were calculated considering not only the superposition of the solutions but also considering separately the influence of the edge load, the fluid pressure and the residual stresses. This way it was possible to study the influence of each parameter and how does it contributes to the final solution of the SIFs.

The ellipse dimensions are: a = 0.52 mm and b = 0.83 mm.

Influence of the Hertzian pressure, fluid pressure inside the crack and residual stresses

For each crack growth angle, the following figures (Figures 6.11 – 6.13) present the influence of each load (Hertzian pressure, fluid pressure and residual stresses).

(a)



(b)

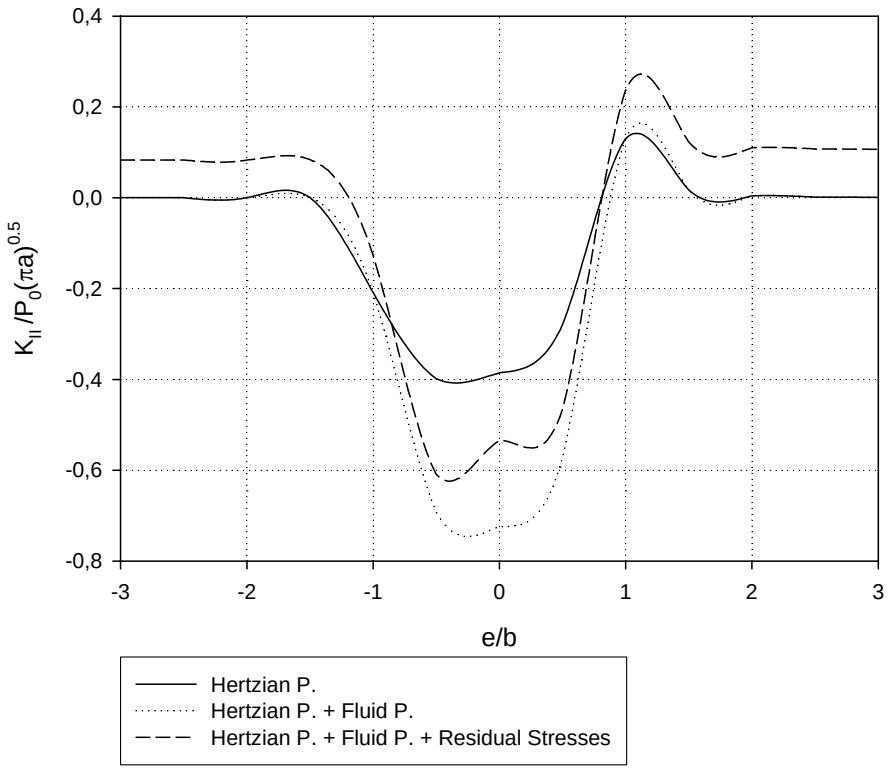
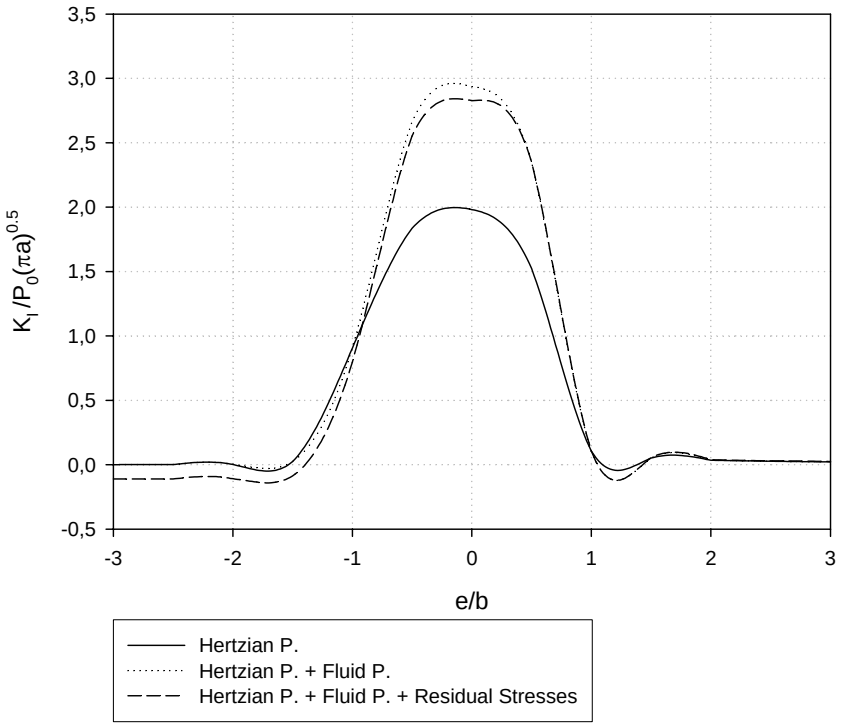


Figure 6.11 – Variations in the non-dimensional SIFs (a) K_I and (b) K_{II} due to movement of the contact for $\theta = 30^o$.

(a)



(b)

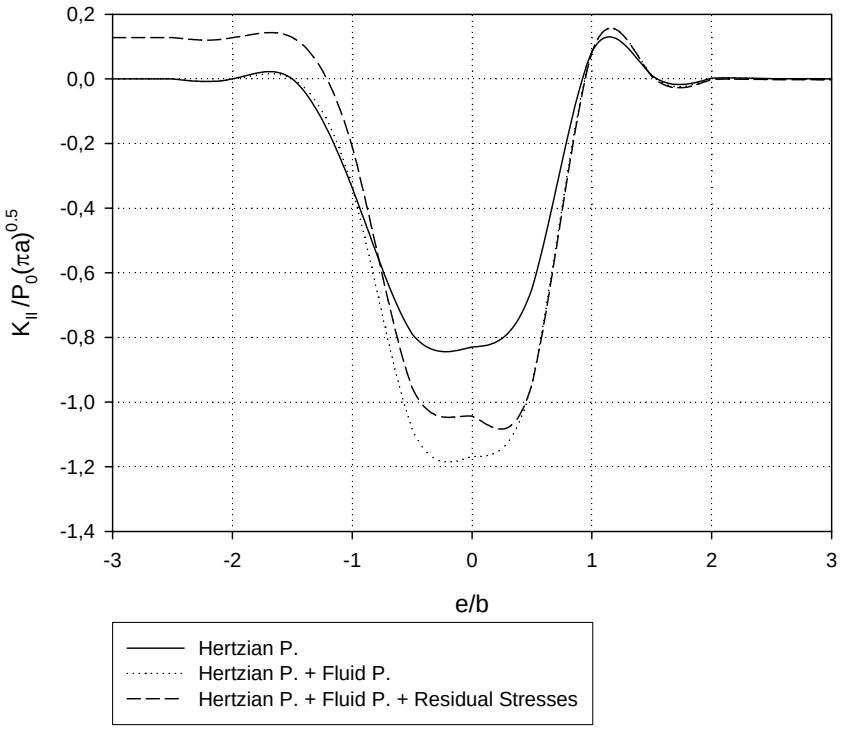
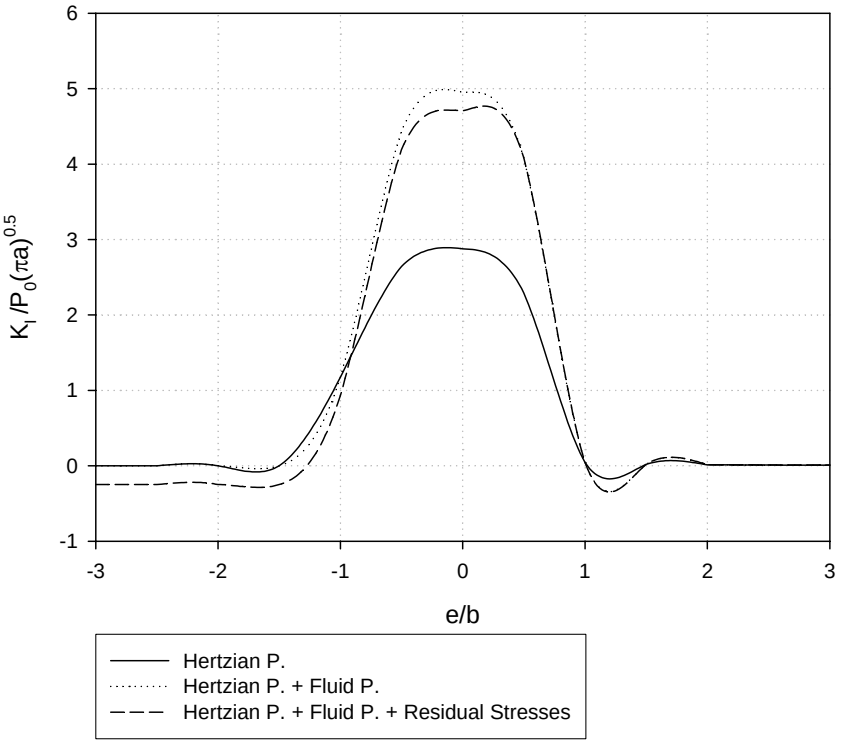


Figure 6.12 – Variations in the non-dimensional SIFs (a) K_I and (b) K_{II} due to movement of the contact for $\theta = 45^o$.

(a)



(b)

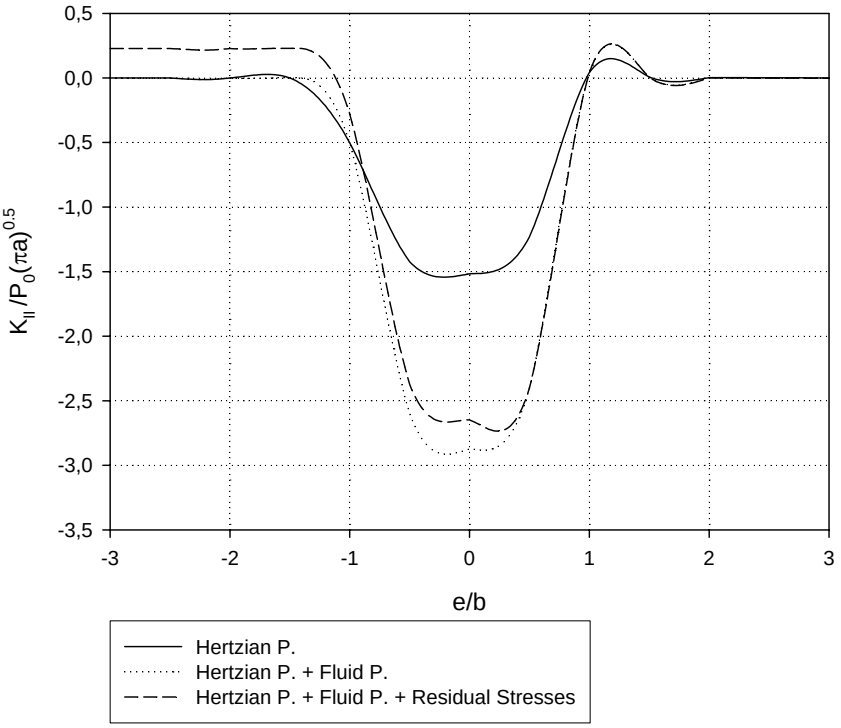
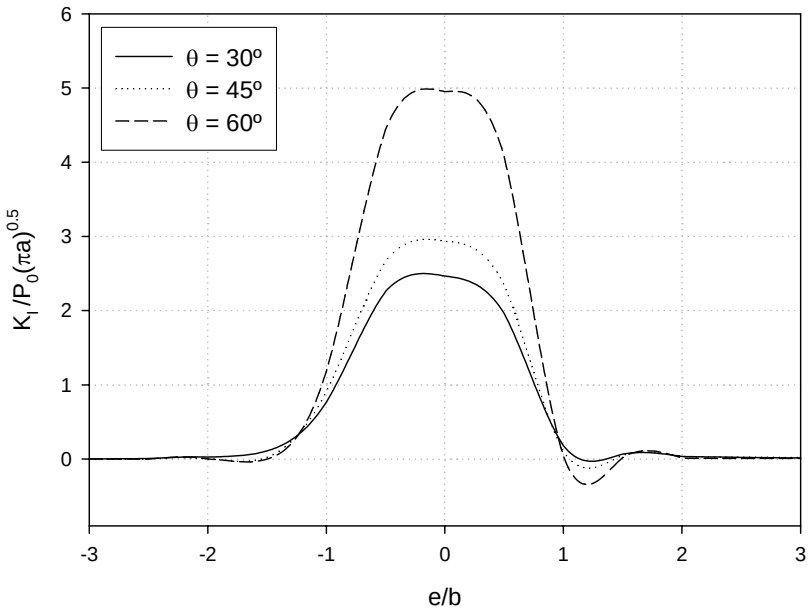


Figure 6.13 – Variations in the non-dimensional SIFs (a) K_I and (b) K_{II} due to movement of the contact for $\theta = 60^o$.

Influence of the crack growth angle

The two figures below (Figure 6.14 and 6.15) compare the influence of the crack growth angle due to the combined influence of the Hertzian pressure + fluid pressure inside the crack and due to the combined influence of these two parameters plus the residual stresses.

(a)



(b)

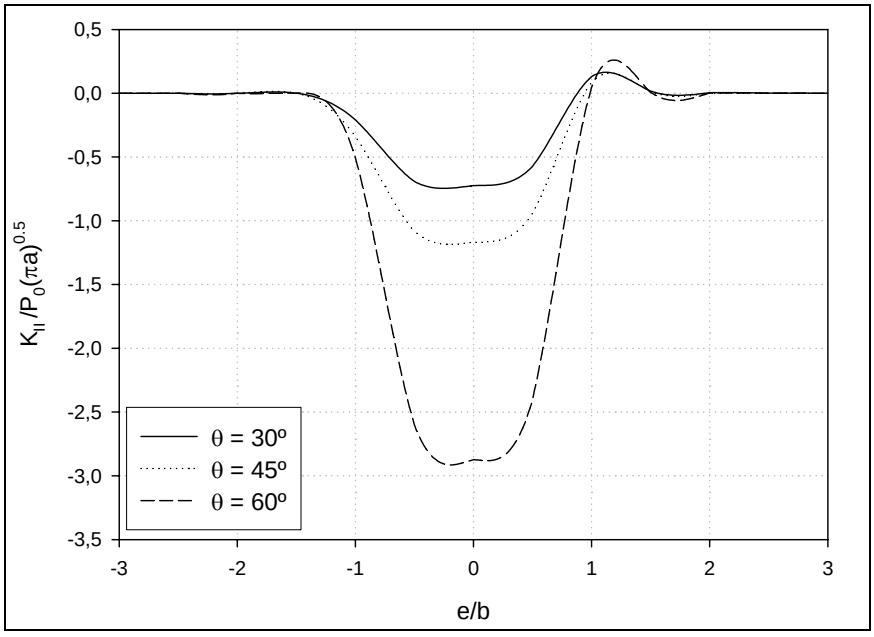
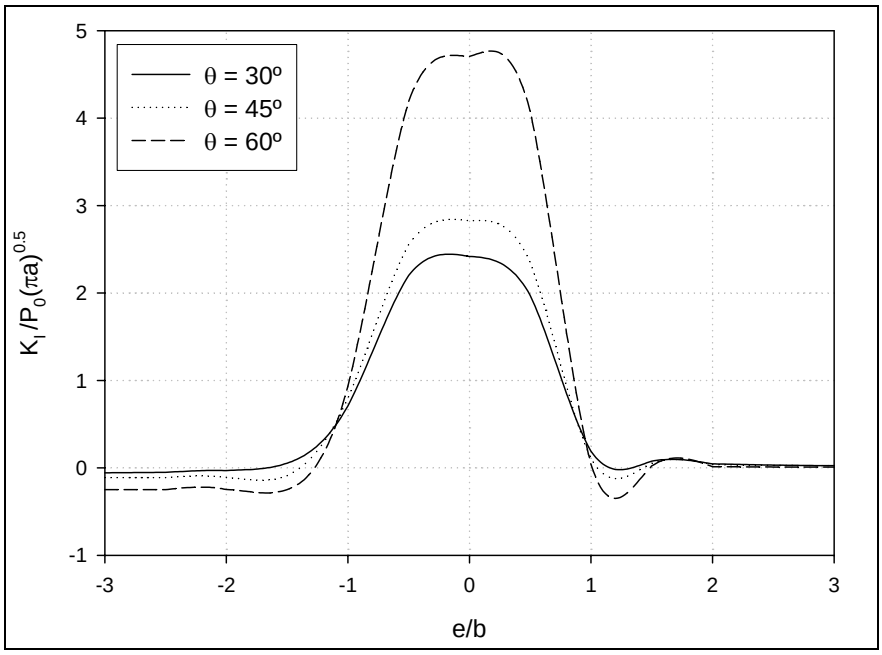


Figure 6.14 – Non-dimensional SIFs (a) K_I and (b) K_{II} for different crack growth angles (θ = 30º, 45º and 60º) considering the influence of the Hertzian pressure and fluid pressure inside the crack.

(a)



(b)

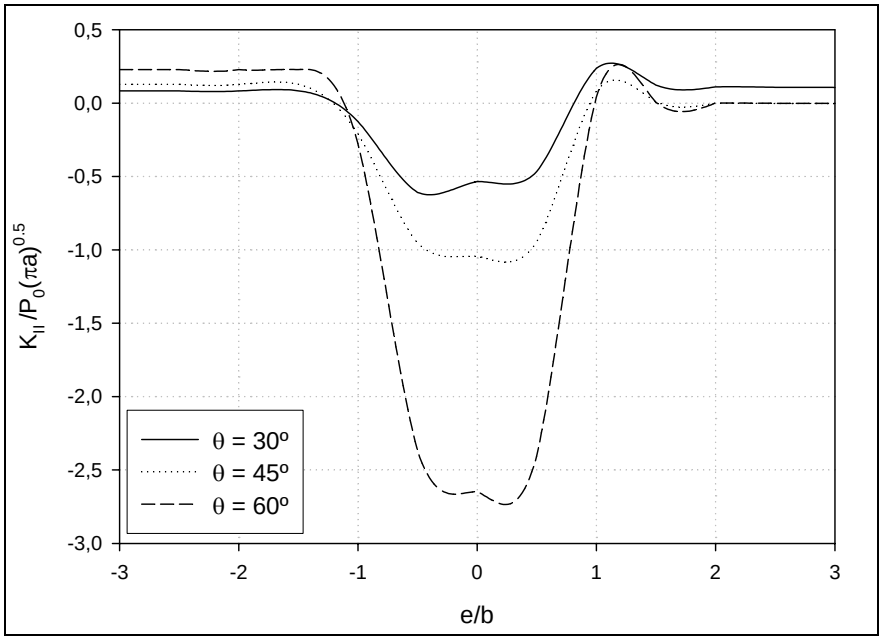


Figure 6.15 – Non-dimensional SIFs (a) K_I and (b) K_{II} for different crack growth angles (θ = 30º, 45º and 60º) considering the influence of the Hertzian pressure, the fluid pressure inside the crack and the residual stresses due to an artificial dent.

Conclusions

A method using the Green's function was used to calculate the SIFs for a surface-breaking crack under the influence of the moving Hertzian pressure, the fluid pressure inside the crack and the residual stresses due to the existence of a dent.

For all the cases, results showed that the higher values of K_I and K_{II} are found when Hertzian pressure is moving closer to the crack, in a range e/b = ± 1, which makes sense because it is in this range that the Hertzian pressure is moving on the crack. It was also found that considering the influence of the fluid pressure inside the crack there is an increase on the values of both K_I and K_{II} in the same range, which also makes sense because in this range a pressure at the crack mouth exists.

Concerning the influence of the residual stresses due to an artificial dent, little changes can be found for K_I. For K_{II} an increase exists when the Hertzian pressure is not moving on the crack and a decrease exists when this pressure is moving on it. It can then be concluded that when the Hertzian pressure is not moving on the crack, the tendency for crack propagation is higher in the presence of residual stresses. On the other hand, when the Hertzian pressure is moving on the crack, the existence of a residual stress field contributes for diminishing the K_{II} values and a smaller tendency for crack propagations exists. However, the increasing effect superposes the diminishing one, and the presence of the residual stresses globally benefits crack propagation, as it was seen experimentally.

Finally, results for different crack growth angles showed that an increase in the angle conduces to higher values of both K_I and K_{II}. When comparing the results for crack growth angle of 60° and 45° the increase is more significant than the one than can be found when the results between an angle of 30° and 45° are compared. These conclusions are coherent with the experimental results obtained – a higher crack growth angle conduced to deeper and larger spall craters (see Chapter IV, Section 3.2).

Final conclusions

The influence of the lubricant, either a liquid like oils or a semi-solid like greases, on rolling contact fatigue is difficult to determine due to the complexety of the mechanisms involved that conduct to the final damage.

The main goal of this work was to investigate the influence of grease composition on surface damage by rolling contact fatigue. To achieve it, experimental tests were carried out in a twin-disc machine. Three greases with different properties and correspondent base oils were tested. The variable properties were the base oil viscosity, the % of soap concentration and the presence or not of additives. To localize and accelarate damage by fatigue a technique using artificial dents was applied. Periodical stops and a video microscope device was used to perform surface observations on the evolution of the damaged area around the dents and on the state of the remaining surface. Pictures were taken and analysed and in the end of the tests it was possible to compare the damaged areas growth around the dents and the wear volume evolution for the different lubricants. Sub-surface observations provided information on crack propagation near the dents and on the profile of spalls at the end of the tests.

From the analysis based on the experimental results concerning the influence of grease composition on the damage by fatigue the final conclusions are that:

- An increase in oil viscosity results in a higher value of λ and consequently in a smaller damaged area around the dents and less wear. This was proven by the tests using the base oils.
- An increase in % of soap concentration benefits the protection of the surfaces to contact fatigue damage. This was proven by the tests using greases G1 (9 % soap concentration) and G2 (12 % soap concentration) that exhibit a similar base oil viscosity but different % of soap concentration. The values of the damaged area around the dents were higher for the tests using G1.
- The existence of an additive, such as EP, in grease G3, ally to a higher viscosity of the base oil, lead to an almost inexistent damaged in the end of the tests. In this case it is believed that the influence of the additive was prevailed on the other two parameters (base oil viscosity and % of soap concentration).

It was found that these conclusions are consistent with the results that were obtained by several authors that studied the influence of grease composition on the film thickness and concluded that under fully-flooded conditions an increase in base oil viscosity and % of soap composition leads to a greater film thickness and consequently to a better protection of the surfaces against contact fatigue.

Concerning the residual stresses it was found that when comparing the measurements before and after the tests (before the tests: σ_{11} = - 420 MPa; after

the tests: σ_{11} = - 1260 MPa) a considerable increase on the residual stresses existed which was probably due to high plastic deformation that occurred in the contacting zone, particularly in the center of the contact, where the Hertzian pressure distribution reaches its maximum value.

To complement the experimental study a numerical analysis was also performed. The indent process used to print artificial dents in the discs surface was simulated using the FEM. Comparison of the final dent profile between experimental measurements and numerical results showed a good agreement, validating the finite element simulation. The results for equivalent plastic strain and residual stresses field after unloading showed that high values of both plastic strain and residual stresses were introduced. It was found that the higher values were situated at the dent's shoulder, which confirms the experimental results because it was precisely at the dent's shoulder that the initial damage by fatigue took place.

Finally, a calculation of the SIFs for a surface-breaking crack under the influence of the moving Hertzian pressure, the fluid pressure inside the crack and the residual stresses due to the existence of a dent was carried out with a method that uses the edge and line Green functions. This calculation was important to quantify the influence of the referred parameters in the final SIFs and also to determine the influence of the crack growth angle.

Results showed that the higher values of K_I and K_{II} are found when Hertzian pressure is moving closer to the crack, in a range e/b = ± 1 because it is in this range that the Hertzian pressure is moving on the crack. It was also found that considering the influence of the fluid pressure inside the crack there is an increase on the values of both K_I and K_{II} because it is in this range that an pressure at the crack mouth exists.

Concerning the influence of the residual stresses due to an artificial dent, little changes were found for K_I. For K_{II} an increase existed when the Hertzian pressure was not moving on the crack and a decrease existed when this pressure was moving on it. It can therefore be concluded that when the Hertzian pressure is not moving on the crack, the tendency for crack propagation is higher in the presence of residual stresses. On the other hand, when the Hertzian pressure is moving on the crack, the existence of a residual stress field contributes for diminishing the K_{II} values and a smaller tendency for crack propagations exists. However, the increasing effect superposes the diminishing one, and the presence of the residual stresses globally benefits crack propagation, as it could be seen experimentally.

Finally, results for different crack growth angles showed that an increase in the angle conduces to higher values of both K_I and K_{II}. When comparing the results

for crack growth angle of 60° and 45° the increase is more significant than the one than can be found when the results between an angle of 30° and 45° are compared. These conclusions are coherent with the obtained experimental results where a higher crack growth angle conduced to deeper and larger spall craters.

Future work

The final conclusions of this work suggest a natural direction for future work. Several tasks both on experimental and numerical investigations can be carried out in a study that continues the one that just finishes.

Experimental investigations

- Perform fatigue tests on the twin-disc machine for a wider range of greases in order to test more accurately the influence of the studied components: base oil viscosity, % of soap composition and additives. Extend this study to incorporate the influence of the type of thickener (lithium, calcium, aluminium, etc.).
- Run the experimental tests not only under pure rolling conditions but also introducing different values of sliding.
- Analyse the influence of the maximum Hertzian pressure by running tests for different values of P_0 (for example 1.5, 2.5 and 3.5 GPa).
- Investigate, if possible, the occurrence of RCF not only under fully-flooded but also under starved conditions.

Numerical investigations:

- Complete the finite element analyses for the indent process by calculating also the contact between the indented surface and a smooth one for the contact loads used in the experimental tests.
- Perform residual stresses measurements in-depth and in both longitudinal and transversal directions for different stages of the fatigue tests, in order to analyse how the residual stress distribution evolutes.
- Concerning the stress intensity factors calculation, introduce the variation of the Hertzian pressure distribution as it passes a dent.
- Extend the two-dimensional analysis to a three-dimensional one, introducing more influent parameters (sliding, temperature, lubricant behaviour inside the crack, etc.), creating a more realistic model.

References

[1] Gohar, R., "Elastohydrodynamics", Ellis Horwood Lid., 1988.

[2] Tallian, T. E., "Rolling Contact Fatigue", SKF Ball Bearing Journal, Vol. 217, pp. 5-13, 1983.

[3] Tallian, T. E., "Failure Atlas for Hertz Contact Machine Elements", ASME Press, 1992.

[4] Harris, T.A., Barnsby, R.M., "Life ratings for ball and roller bearings", Proc. Instn. Mech. Engrs., Vol. 215, Part J, pp. 577-595, 2001.

[5] Hertz, H., "Über die berührung fester elastischer körper". J Reine und Angew, Math., Vol. 92, pp. 156-171, 1881.

[6] Dumont, M.-L., "Etude des endommagements de surface induits par la fatigue de roulement dans les contacts élastohydrodynamiques pour des aciers M50 et 100Cr6", PhD thesis, INSA, Lyon, France, 1997.

[7] Way, S., "Pitting Due to Rolling Contact". ASME Trans., J. Appl. Mech., Vol.2, pp. A49-A58, 1935.

[8] Lundberg, G., Palmgren, A., "Dynamic Capacity of Rolling Bearings". London: Acta Polytechnica, Mechanical Engineering Series, Vol. 1, nº3, 1947.

[9] Ioannides, E., Harris, T.A., "A New Fatigue Life Model for Rolling Bearings". ASME Trans., J. Trib., Vol. 107, nº 3, p. 367-378., 1985.

[10] Stachowiak, G. W., Batchelor, A. W., "Engineering Tribology", Elsevier Science Publishers B. V., 1993.

[11] Johnson, K.L., "Contact Mechanics", Cambridge University Press, 1985.

[12] Alfredsson, B. "A study on contact fatigue mechanisms", Doctoral thesis, Department of Solid Mechanics, Royal Institute of Technology, Stockholm, Sweden, 2000.

[13] Hamrock, Dowson "Ball Bearing Lubrication, The Elastohydrodynamics of Elliptical Contacts". John Willey & Sons, 1981.

[14] Tallian, T.E., "On Competing Failure Modes in Rolling Contact", *ASLE Transactions*, Vol. 10, pp. 418-439, 1967.

[15] Seabra, J., Campos, A., Sottomayor, A., "Lubrificação elastohidrodinâmica", FEUP, 2002.

[16] Tallian, T.E., "A unified model for rolling contact life prediction". Trans. ASME, JOLT, Vol. 104, pp. 336-346, 1982.

[17] Voskamp, A.P., "Material response to rolling contact loading". Trans. ASME, JOT Vol. 107, pp. 359-366, 1985.

[18] Martin, J.A., Borgese, S.F., Eberhardt, A.D., "Micro-structural alterations of rolling bearing steel undergoing cyclic stressing". Trans. ASME, J. Basic Eng., Vol. 88, pp. 555-567, 1966.

[19] Bower, A.F., "The influence of crack face friction and trapped fluid on surface initiated rolling contact fatigue cracks", Trans. ASME, JOT, Vol. 110, pp. 704-711, 1988.

[20] Bhargava, V., Hahn, G.T., Rubin, C.A., "Rolling contact deformation etching effects and failure of high-strength bearing steel", Metallur. Trans. A, Vol. 21A, pp. 1921-1931, 1990.

[21] Suh, N.P., "An overview of the delamination theory of wear", WEAR, Vol.44, Nº1, pp. 1-16, 1977.

[22] Fan, H., Keer, L. M., Cheng, W., Cheng H. S., "Competition Between Fatigue Crack Propagation and Wear", J. Trib., Vol. 115, pp. 141-147, 1993.

[23] Cheng, W., Cheng, H.S., Keer, L.M., "Longitudinal crack initiation under pure rolling contact fatigue", STLE Trib. Trans., Vol. 37, nº1, pp. 51-58, 1994.

[24] Cheng, W., Cheng, H.S., Mura, T., Keer, L.M., "Micromechanics Modeling of Crack Initiation Under Contact Fatigue", ASME Trans., J. Trib., Vol. 116, nº1, pp. 2-8, 1994.

[25] Clarke, T.R., Miller, G.R., Keer, L.M., Cheng, H.S., "The role of near-surface inclusions in the pitting of gears". ASLE Trans., Vol. 28, pp. 111-116, 1985.

[26] Cheng, W., Cheng, H.S., “Effect of surface roughness orientation on pitting resistance of lubricated rollers”. STLE, Trib. Trans., Vol. 38, pp. 396-402, 1995.

[27] Tallian, T.E., “Simplified contact fatigue life prediction model – Part I: Review of published models”, Journal of Tribology, Vol. 114, pp. 207-213, 1992.

[28] Tallian, T.E., “Simplified contact fatigue life prediction model – Part II: New model”, Journal of Tribology, Vol. 114, pp. 214-222, 1992.

[29] Bold, P. E., Brown, M. W., Allen, R. J., “A Review of Fatigue Crack Growth in Steels Under Mixed Mode I and II Loading”, Fatigue Fract. Eng. Mtrl. Struct., Vol. 15, pp. 965-977, 1992.

[30] Qian J, Fatemi A. “Mixed Mode Fatigue Crack Growth: A Literature Survey”, Eng. Fract. Mech., Vol. 55, pp. 969-990, 1996.

[31] Mura, T., Nakasone, Y., “A theory of fatigue crack initiation in solids”, ASME Trans., J. Appl. Mech., Vol. 57, pp. 1-6, 1990.

[32] Zhou, R.S., Cheng, H.S., Mura, T., “Micropitting in rolling and sliding contact under mixed lubrication”, ASME Trans., J. Trib., Vol. 111, nº4, pp. 605-613, 1989.

[33] Yamashita, N., Mura, T., Cheng, H.S., “Effect of stresses induced by a spherical asperity on surface pitting in elastohydrodynamic contacts”, ASLE Trans., Vol. 28, pp. 11-20, 1985.

[34] Keer, L.M., Bryant, M.D., “A pitting model for rolling contact fatigue”, ASME Trans., J. Lubr. Tech., Vol. 105, n°2, pp. 198-205, 1983.

[35] Hanson, M.T., Keer, L.M., “An analytical life prediction model for the crack propagation occurring in contact fatigue failure”, STLE Tribology Trans., Vol. 35, n°3, pp. 451-461, 1992.

[36] Kaneta, M., Yatsuzuka, H., Murakami, Y., “Mechanism of crack growth in lubricated rolling/sliding contact”, ASLE Trans., Vol. 28, pp. 407-414, 1985.

[37] Kaneta, M., Murakami,Y., “Effects of oil hydraulic pressure on surface crack growth in lubricated rolling/sliding contact”, Trib. Int., Vol. 20, pp. 210-217, 1987.

[38] Fletcher, D.I., Beynon, J.H., "A simple method of stress intensity factor calculation for inclined fluid filled surface-breaking cracks under contact loading", Proc. Inst. Mech. Engrs., Vol. 213, Part J, pp. 299-304, 1999.

[39] Kaneta, M., Murakami, Y., "Propagation of semi-elliptical surface cracks in lubricated rolling/sliding elliptical contacts", ASME Trans., J. Trib., Vol. 113, nº2, pp. 270-275, 1991.

[40] Murakami, Y., Sakae, C., Ichimaru, K., "Three-dimensional fracture mechanics analysis of pit formation mechanism under lubricated rolling-sliding contact loading", STLE Trib. Trans., Vol. 37, nº3, pp. 445-454, 1994.

[41] Dang-Van, K., "Macro-micro approach in high-cycle multiaxial fatigue", Advances in multiaxial fatigue, ASTM STP 1191. D.L. McDowell and R. Ellis, Eds., American Society for Testing and Materials, Philadelphia, pp. 120-130, 1993.

[42] Ekberg, A., Bjarnehed, H., "Rolling Contact Fatigue of Wheel/Rail Systems - A Literature Survey", Report F209, Department of Solid Mechanics, Chalmers University of Technology, 52 pp., 1998.

[43] Cheng, W., Cheng, H.S., Keer, L.M., Ai, X., "Surface crack initiation under contact fatigue: experimental observation and contact analysis, ASME Trans., Vol.115, pp. 658-665, 1993.

[44] Cheng, W., Cheng, H.S., Keer, L.M., "Experimental investigation on rolling/sliding contact fatigue crack initiation with artificial defects, Tribol. Trans., Vol.37, pp. 1-12, 1994.

[45] Nélias, D., Dumont, M.-L., Couhier, F., Dudragne, G., Girodin, D., "Experimental and theoretical investigation on rolling contact fatigue of AISI 52100 and M50 steels under EHL o Micro-EHL conditions", ASME J. Tribol., Vol.124, pp. 653-667, 1997.

[46] Nélias, D., Ville, F., "Detrimental effects of debris dents on rolling contact fatigue", ASME J. Tribol., Vol.122, pp. 55-64, 2000.

[47] Magalhães, L., Seabra, J., Sá, C., "Experimental observations of contact fatigue crack mechanisms for austempered ductile iron (ADI) discs", Wear, Vol. 246, pp. 134-148, 2000.

[48] Girodin, D., Ville, F., Guers, R., Dudragne, G., "Rolling contact fatigue tests to investigate surface initiated damage using surface dents", Bearing Steel Technology, ASTM STP 1419, J.M. Beswick, Ed., ASTM, pp. 1-19, 2002.

[49] Magalhães, L., "Caracterização tribológica de um ferro nodular austemperado em ensaios disco-disco e de engrenagens FZG", PhD thesis (in Portuguese), FEUP, 2003.

[50] Diab, Y., Coulon, S., Ville, F., Flamand, L., "Experimental investigations on rolling contact fatigue of dented surfaces using artificial defects: subsurface analyses", Tribological Research and Design for Enginneering Systems, D. Dowson et al. (Editors), pp. 359-366, 2003.

[51] Nélias, D., Jacq, C., Lormand, G., Dufragne, G., Vincent, A., "New methodology to evaluate the rolling contact fatigue performance of bearing steels with surface dents: application to 32CrMoV13 (Nitrided) and M50 steels", ASME J. Tribol., Vol.127, pp. 611-622, 2005.

[52] Ko, C.N., Ioannides, E., "Debris denting – the associated residual stresses and their effect on the fatigue life of rolling bearings: an FEM analysis", Proceedings of 15th Leeds-Lyon Symposium on Tribology, Leeds, England, 1988, D. Dowson et al. (Editors), pp. 199-207, 1989.

[53] Xu, G., Sadeghi, F., "Spall Initiation and Propagation Due to Debris Denting," Wear, Vol.201, pp. 106–116, 1996.

[54] Xu, G., Sadeghi, F., Cogdell, J.D., "Debris Denting Effects on Elastohydrodynamic Lubricated Contacts," ASME J. Tribol., Vol. 119, pp. 579-587, 1997.

[55] Xu, G., Sadeghi, F., Hoeprich, M.R., "Residual Stresses Due to Debris Effects in EHL Contacts," Tribol. Trans., Vol. 40, pp. 613-620, 1997.

[56] Zhao, J., Sadeghi, F., Nixon, H.M., "A Finite Element Analysis of Surface Pocket Effects in Hertzian Line Contact," ASME J. Tribol., Vol. 122, pp. 47-54, 2000.

[57] Kang, Y.S., Sadeghi, F., Hoeprich, M., "A Finite Element Model for Spherical Debris Denting in Heavily Loaded Contacts," ASME J. Tribol., Vol. 126, pp.71-80, 2004.

[58] Howell, M.B., Rubin, C.A., and Hahn, G.T., "The Effect of Dent Size on the Pressure Distribution and Failure Location in Dry Point Frictionless Rolling Contacts," ASME J. Tribol., Vol. 126, pp. 413-421, 2004.

[59] Coulon, S., Jubault, I., Lubrecht, A.A., Ville, F., Vergne, P., "Pressure profiles measured within lubricated contacts in presence of dented

surfaces. Comparison with numerical models", Tribology International, Vol. 37, pp.111-117, 2004.

[60] Coulon, S., Ville, F., Nélias, D., "Numerical and experimental investigations on rolling contact fatigue for dented surfaces", Proceedings of the 27th Leeds-Lyon Simposium on Tribology, Lyon Sept. 2000, G. Dalmaz & al. Editors, Amsterdam: Elsevier, pp. 459-467, 2001.

[61] Ville, F., Nélias, D., "Early fatigue failure due to dents in EHL contacts", ALSE Transactions, Vol. 42, pp. 795-800, 1999.

[62] Ai, X., Nixon, H.P., "Fatigue life reduction of roller bearings due to debris denting: Part II – Experimental validation", Trib. Trans., Vol. 43, pp. 311-317, 2000.

[63] Coulon, S., Ville, Lubrecht, A.A., "An abacus for predicting the rolling contact fatigue life reduction due to debris dens", Proceedings of the 28th Leeds-Lyon Symposium on Tribology, D. Dowson, C.M. Taylor and M. Godet, Elsevier, Amsterdam, in Tribology Series, Vol. 40, pp. 283-294, 2002.

[64] Lubrecht, A.A., Dwyer-Joyce, R.S., Ioannides, E., "Analysis of the influence of indentations on contact life", Proceedings of 19th Leeds-Lyon Symposium on Tribology, D. Dowson, C.M. Taylor and M. Godet, Elsevier, Amsterdam, pp. 173-181, 1993.

[65] Sayles, R.S., Ioannides, E., "Debris damage in rolling bearings and its effects on fatigue life", Journal of Tribology, Vol. 110, pp. 26-31, 1988.

[66] Cann, P.M.E., "Grease Lubrication of Rolling Element Bearings – Fundamental Mechanisms and Predictive Models", Séminaire commun, LaMCoS-Ecole Doctorale MEGA, 2004.

[67] Cann, P.M.E., "Understanding grease lubrication", The Third Body Concept / D. Dowson et al. (Editors), pp. 573-581, 1996.

[68] Couronné, I., Vergne, P., Mazuyer, D., Truong-Dinh, N., Girodin, D., "Effects of grease composition and structure on film thickness in rolling contact", Trib. Trans., Vol. 46, 1, pp. 31-36, 2003.

[69] Miettinen, J., Andersson, P., Wikström, V., "Analysis of grease lubrication of a ball bearing using acoustic emission measurement", Proc. Instn Mech. Engrs, Vol. 215, Part J, pp. 535-544, 2001.

[70] Cann, P., "Starvation and reflow in a grease lubricated elastohydrodynamic contact", Tribology Trans., Vol. 39, pp. 698-704, 1996.

[71] Jonkisz, W., Krzeminski-Freda, H., "The properties of elastohydrodynamic grease films", Wear, Vol. 77, pp. 277-285, 1982.

[72] Aihara, S., Downson, D., "A study of film thickness in grease lubricated elastohydrodynamic contacts", Proceedings of the 5th Leeds-Lyon Symposium on Tribology, Leeds, pp. 104-115, 1979.

[73] Åström, H., Isaksson, O., Höglund, E., "Video recordings of an EHD point contact lubricated with grease", Tribology Int., Vol. 24, pp. 179-184, 1991.

[74] Palacios, J.M., Cameron, A., Arizmendi, L., "Film thickness of grease in rolling contacts", ASLE Trans., Vol. 24, pp. 474-478, 1981.

[75] Kendall, D.R., Williamson, B.P., "The influence of grease composition on film thickness in EHD contacts", ELGI Conference, pp. 28-36, 1992.

[76] Kaneta, M., Ogata, T., Tabuko, Y., Naka, M., "Effects of a thickener structure on grease elastohydrodynamic lubrication films", Proc. I. Mech. E., Vol. 214, Part J, pp. 327-336, 2000.

[77] Hurley, S., Cann, P.M., "Grease composition and film thickness in rolling contacts", Eurogrease, pp. 7-25, 1999.

[78] Cann, P., Lubrecht, A.A., "An analysis of the mechanisms of grease lubrication in rolling element bearings", Lubrication Science, Vol. 11, pp. 227-245, 1999.

[79] Cann, P.M.E., "Starved grease lubrication of rolling contacts", Tribology Transactions, Vol. 42, pp. 867-873, 1999.

[80] SAE Standard, Automotive Lubricating Greases, SAE J310, 1987.

[81] Amaro R.I., Martins R.C., Seabra J.O., Brito A.T., "MoST low friction coating for gears application", Gears & Transmissions Workshop, FEUP, pp. 206-207, 2003.

[82] Kandil, F.A., Lord, J.D., Fry, A.T., Grant, P.V., "A review of residual stress measurement methods – a guide to technique selection", NPL Report MATC(A)04, 2001.

[83] Withers, P.J., Bhadeshia, H.K.D.H., "Residual stress. Part 1 – Measurement techniques", Material Science and Technology, Vol. 17, pp. 355-365, 2001.

[84] "Residual stress theory" in www.physiqueindustrie.com, Physique & Industrie, 2003.

[85] Rooke D.P., Rayaprolu D.B., Aliabadi M.H., "Crack-line and edge Green's functions for stress intensity factors of inclined edge cracks", *Fatigue Fracture Engng Mater. Structs*, Vol. 15, pp. 441-461 (1992).

[86] Fletcher, D.I., Beynon, J.H., "A Simple Method of Stress Intensity Factor Calculation for Inclined Surface-Breaking Cracks with Crack Face Friction under Contact Loading", Proc. Inst. Mech. Engrs., Vol. 213, Part J, pp. 481-486(6), 1999.

[87] Tyfour, W. R., Beynon, J. H., "The effect of rolling direction reversal on fatigue crack morphology and propagation", Tribology Int., Vol. 27, pp. 401-412, 1994.

[88] Benuzzi D., Bormetti E., Donzella G., "Modelli numerici per lo studio della propagazione di criche superficiali da rolling contact fatigue in presenza di fluido", XXX Convegno Nazionale AIAS – Alghero (SS), pp. 12-15, 2001.

Appendix A

Re-design of the twin-disc machine

The twin-disc machine was originally design for tests using oils. In order to test greases, a re-design had to be made, particularly in the body that contains the two contacting discs (discs box).

Figure A1 shows a detailed drawing of one of the discs box from the original twin-disc machine design. The oil enters the contact to lubricate the rotating discs and also the roller bearings that support the shaft where the disc and the pulley is inserted (for a better understanding see also Chapter IV, Section 2.2).

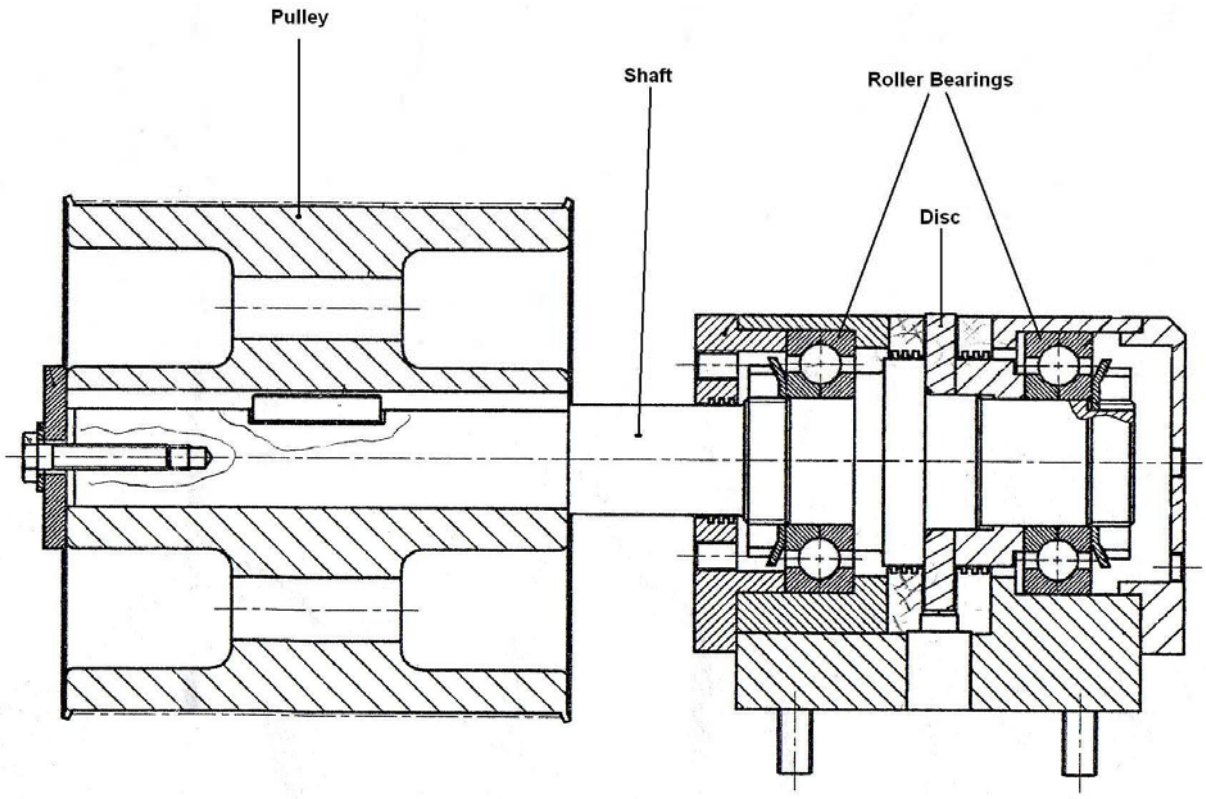


Figure A1 – Detailed drawing of the discs box for the original twin-disc machine design.

For the tests using greases, the discs were evidently lubricated with grease, but the roller bearings had to be lubricated with oil, to prevent eventual failures. Therefore, a re-design was performed to make possible an independent lubrication supply for the discs and roller bearings. Figure A2 shows the changes that were made.

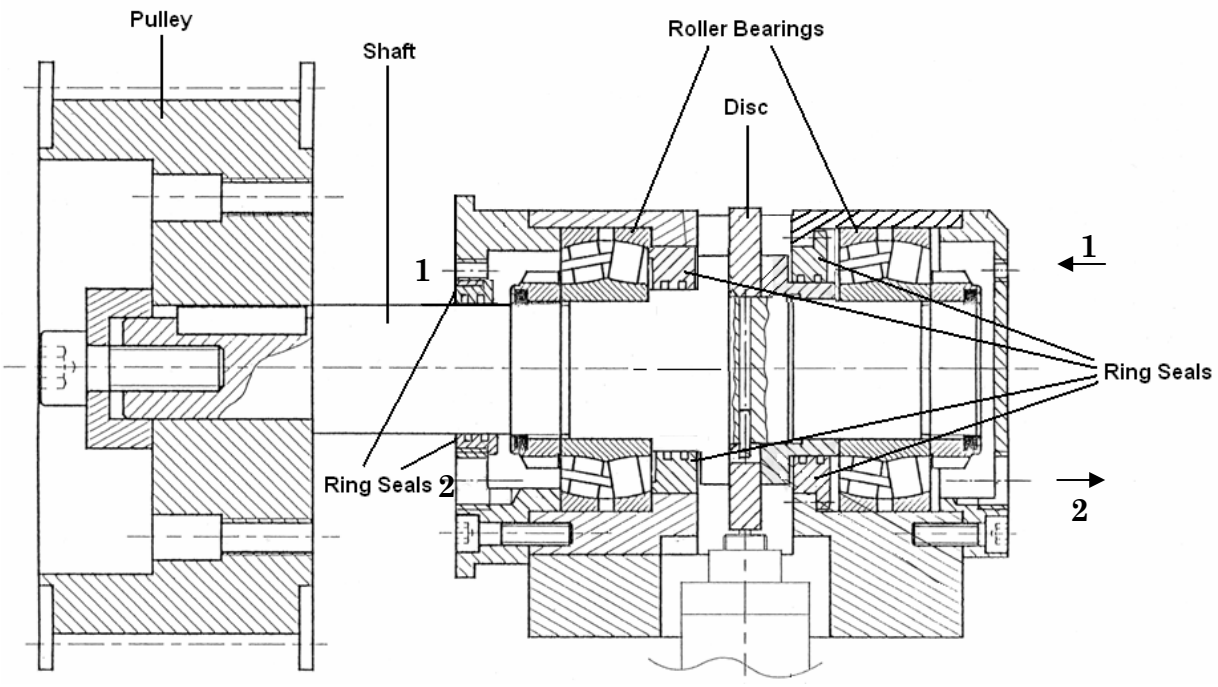


Figure A2 – Detailed drawing of the discs box for the modified twin-disc machine design.

The oil stored in a tank travels through two pipes entering (1) the space occupied by the roller bearings and leaving (2) from two other pipes that return to the tank.

Before running the tests, grease is manually placed in two parts that cover the discs. These parts store an amount of grease that will be replenish to the rolling contact providing fully-flooded lubrication conditions.

To ensure that the oil used to lubricate the roller bearings will not get into the contact, several ring seals were installed in the interface that separates the roller bearings and the zone were the contact is being grease lubricated.